CHEMICAL AND CONFORMATIONAL STUDIES OF BACTERIAL CELL SURFACE POLYSACCHARIDE REPEATING UNITS

Zaheer Timol

Supervisors: Neil Ravenscroft, Michelle Kuttel and David Gammon

A thesis presented for the degree of
Master in Science

University of Cape Town
March 2017

Abstract

Bacterial cell surface polysaccharides are primarily present as lipopolysaccharides or capsular polysaccharides. They are used by cells for both structure and function and have been shown to be a virulence factor of bacterial pathogens. Cell surface polysaccharides are widely utilised as antigenic components in vaccines and play an important role in the protection against numerous diseases including meningococcal disease and shigellosis. This study is composed of two parts: a computational section, which investigates the capsular polysaccharide (CPS) repeating unit (RU) conformations of meningococcal Y and W CPS vaccines and a second experimental component that involves synthetic studies toward the O-specific polysaccharide (O-SP) RU of *Shigella sonnei*.

The CPS RU of MenY [→6)-α-D-Glc(1→4)-α-D-NeuNAc-(2→] and MenW [→6)-α-D-Gal(1→4)-α-D-NeuNAc-(2→] differ only in the orientation of the C-4 hydroxyl: equatorial in MenY and axial in MenW. However, groups Y and W CPS vaccines have different levels of antibody cross-protection. The purpose of the computational study was to determine if these observed differences may be attributed to CPS RU conformation.

Potential of mean force calculations were applied to disaccharide RUs of MenY and MenW, and larger three repeating units (3RU) were simulated with molecular dynamics (MD) in solution. The molecular modelling showed that differences in RU conformation between the meningococcal groups arise primarily due to the structural differences between glucose and galactose; affecting the behaviour and orientation of the 2→6 dihedral linkage. The 2→6 linkages in the MenY 3RU adopt a single preferred orientation and consequently it has a single dominant molecular conformation. In contrast, the 2→6 linkages in the MenW 3RU move frequently between different rotameric conformations resulting in multiple conformational families. These results indicate significant conformational differences between the MenY and MenW CPS RUs, which may account for the different levels of cross-protection observed.

The synthetic component was part of a larger study to develop a novel route towards the O-SP RU of *S. sonnei* for use in biological testing and physicochemical characterisation for vaccine development. The O-SP RU of *S. sonnei* is →4-α-L-AltNAc-(1→3)-β-D-FucNAc4N(1→. The multi-step synthesis was performed using known methodology as well as methods developed by the research group. The key 2,3-oxazolidinone protected intermediate was successfully synthesised in good yields and due to time constraints the final product synthesised was two steps away from the protected FucNAc4N residue.

Additional studies were performed on the 2,3-oxazolidinone intermediate as part of a divergent synthesis strategy toward the AltNAcA residue of *S. sonnei*. Reactions were conducted whereby β and α derivatives of the 2,3-oxazolidinone intermediate were successfully synthesised in large scale and good yields for further studies to be performed by the group.

Acknowledgments

I wish to acknowledge the contribution of the following people without whom this project would not have been possible:

First, I would like to thank my supervisors - Associate Professor's Neil Ravenscroft, Michelle Kuttel and David Gammon for their advice, support and considerable patience throughout the project.

To all my friends and family members for their continued encouragement.

Students from the Gammon, Hunter and Chibale Research Groups.

The University of Cape Town and the National Research Foundation for funding support.

The University of Cape Town's ICTS High Performance Computing team for providing the computation facilities: http://hpc.uct.ac.za. Many thanks to Andrew Lewis for all his assistance.

Contents

List of Figures

List of Tables

List of Schemes

Abbreviations

AltNAcA	2-acetamido-2-deoxy-L-altruronic acid
Boc	Di-tertiarybutyl dicarboxalate
BCR	B-cell receptor
CAS	Ceric ammonium sulfate
CPS	Capsular polysaccharide
DFT	Density functional theory
dOA	de-O-acetylated
FES	Free energy surface
FucNAc4N	2-acetamido-4-amino-2,4,6-trideoxy-D-galactopyranose
H-Bond	Hydrogen bond
LPS	Lipopolysaccharide
MenA	Meningococcal group A
MenB	Meningococcal group B
MenC	Meningococcal group C
MenW	Meningococcal group W
MenX	Meningococcal group X
MenY	Meningococcal group Y
MD	Molecular dynamic(s)
NAMD	Nano scale molecular dynamics
N. meningitidis	*Neisseria meningitidis*
NMR	Nuclear magnetic resonance
NPT	Constant number of particles, pressure, and temperature
NVT	Constant Number of particles, volume, and temperature
OAc	O-acetylated
O-SP	O-Specific polysaccharide
PBC	Periodic boundary conditions
PMF	Potential of mean force
SASA	Solvent accessible surface area
$\mathbf{R_g}$	Radius of gyration
RU	Repeating unit
RMSF	Root mean square fluctuation
RMSD	Root mean square deviation
S. sonnei	*Shigella sonnei*
sIg	surface immunoglobulin
$\mathbf{T}_H$	Helper T-cells
$\mathbf{T}_I$**-1**	Type-1 independent T-cells
$\mathbf{T}_I$**-2**	Type-2 independent T-cells
VMD	Visual molecular dynamics

Chapter 1. Introduction

The components that surround the cell and form the outermost layers are primarily constituted of proteins, lipids and carbohydrates. They are involved in a wide range of roles including cell protection, adhesion and recognition. Carbohydrates, in particular, are arguably the most abundant and structurally diverse class of organic molecules in nature and, thus, play a pivotal role in living organisms. Therefore, the study of these fundamental molecules and their role in biological processes is of importance.

This chapter provides an overview of bacterial extracellular polysaccharides and the role of carbohydrate-based vaccines in protection against encapsulated organisms.

1.1 Cell surface polysaccharides

Bacteria, in comparison with many cells of higher organisms, are often found in more unpredictable and hostile environments.[1,2] To survive and grow in these harsh conditions they have developed numerous unique cellular traits. One of these is their distinct cell envelope, which comprises of several layers of material that surround and protect the cytoplasm as well as facilitate a wide range of cellular functions.[1] These range from providing structural integrity to the cell, to the selective passage of nutrients and waste, as well as regulating interactions between the cell and its external environment.[1–3]

The cell envelope is a descriptive term for the cellular surface components that form the plasma membrane, cell wall and capsule. Bacterial cells often contain all three components, however, gram-positive bacterium consist of a single plasma membrane whilst gram-negative bacterium contain both an inner and outer membrane. The cell wall and capsule primarily constitute the surface layers with the capsule usually present outermost of the cell envelope. Figure 1.1 shows a cross sectional view of the inner and outer layers of a general gram-negative bacterial cell envelope.

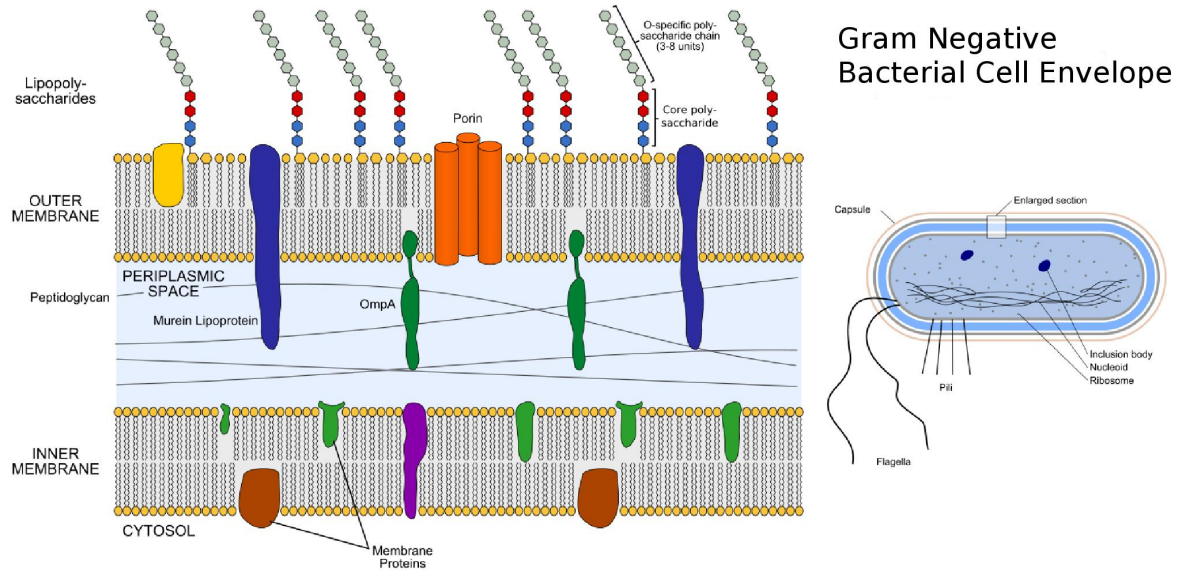


Figure 1.1: Cross-sectional view of a gram-negative cell envelope.[4]

Bacterial cell surface polysaccharides are primarily present as a component of lipopolysaccharide (LPS) or as polysaccharide units in the capsule. LPSs are only present in gram-negative bacteria as a constituent of the outer membrane within which the lipid part of the LPS molecule is embedded.[3] The carbohydrate units of a LPS molecule linked to the lipid tail are present as a core oligosaccharide or an O-specific polysaccharide (O-SP). The O-SP lies facing outward of the cell and is the key biochemical component of LPS.[3]

The bacterial capsule is formed by expression of high molecular-weight cell surface polysaccharides that create a multilayered, amorphous coat surrounding the cell. They can be classified into two main groups: Polysaccharide molecules that loosely associate with the cell surface are regarded as slime polysaccharide whilst those that closely interact with the cell surface are termed capsular polysaccharide (CPS).[5] Bacterial CPSs are generally attached to the cell surface by covalent linkage to a lipid moiety located on the outer membrane (Figure 1.2).[6,7] These can be regarded as being LPS like.

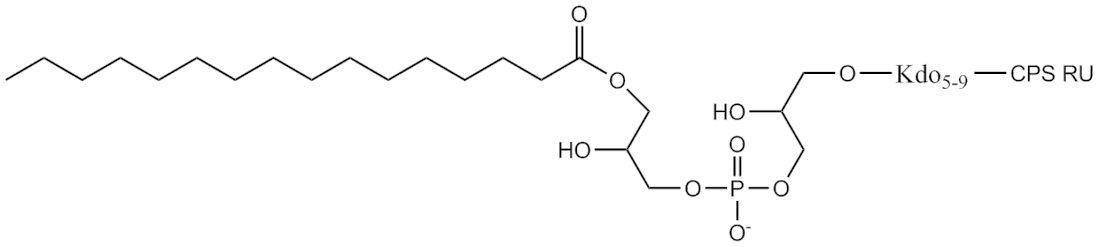


Figure 1.2: Proposed structure of the conserved reducing terminal glycolipid in *E. coli* and *N. meningitidis* CPS.[8]

The long chain polysaccharide molecules that constitute bacterial cell surface polysaccharides are formed by monosaccharides glycosidically linked together in fragments termed a repeating unit (RU). The types of monosaccharides present in bacterial cell surface polysaccharides are wide ranging and can occur in both the pyranose and furanose forms. Furthermore, the glycosidic linkage between any two monosaccharides can bond at any of the hydroxyl groups present on the ring, occur in both α- or β- anomeric forms as well as contain substituents. In addition, organic and inorganic moieties can be part of the primary structure, substituted at certain positions along the polysaccharide or also be introduced as branches alongside the primary chain.[9]. These characteristics alone lead to a high degree of diversity in the primary chemical structure* of cell surface polysaccharide molecules that not only differ in the type of carbohydrate residues but also their linkage positions.

Cell surface polysaccharides are commonly around 200 residues in size, range in weight from 100 to 1000 kDa and can extend approximately 100-400 nm from the cell surface.[10–12] Cell surface polysaccharides are also typically negatively charged due to uronic acids, phosphates or pyruvic acid residues. They also often contain a large number of polar functional groups, leading them to be highly hydrated, with their structures containing more than 95% water.[13]

The chemical composition of O-SPs and CPSs have been extensively researched and are generally well defined for a large number of bacterial types.[1–3] In comparison, molecular conformation of the O-SP and CPS has not been as widely studied. Early conformational studies on the LPS of *Escherichia coli* and

* The primary chemical structure of a cell surface polysaccharide molecule in this text refers to the linear arrangement of monosaccharides in the chain, similar to that of the primary structure of a protein

Salmonella typhimurium were performed by Kastowsky *et al*. More recent work has included further study on *S. typhimurium* as well as *Shigella flexneri* and *Streptococcus pneumoniae*. These studies found that O-SPs and CPSs can be conformationally flexible as well as highly rigid and that small changes in chemical composition can significantly influence the conformational structure and function of cell surface polysaccharides.[14–16]

Bacterial cell surface polysaccharides have been shown to play a role in mediating interactions between a bacterium and its immediate environment and has been implicated as an important factor in the virulence of many animal and plant pathogens.[9,17,18] They have been found to protect cells from perennial effects of desiccation and can also act as storage reserves of carbohydrates for subsequent use in metabolism. Table 1.1 provides some of the key functions of cell surface polysaccharides.

Table 1.1: Functions of bacterial extracellular polysaccharides

Function	**Purpose**
Prevention of dessication	Transmission and survival
Adherence	Colonisation of oral surfaces Colonisation of indwelling catheters Bacteria plant interactions
Resistance to non-specific host immunity	Complement-mediated phagocytosis Complement-mediated killing
Resistance to specific host immunity	Poor antibody response to the capsule

A single bacterial species can produce a range of structurally diverse cell surface polysaccharides and these are frequently the basis of serotyping schemes. However, different bacteria may also produce cell surface polysaccharides that are chemically identical to other bacterial species or other organisms, such as humans. This can make identification and effective treatment particularly challenging.

As cell surface polysaccharides are associated with the surface of pathogenic bacteria they are most often the first antigenic components encountered by the immune system upon infection. They also have a protective function as bacteria that do not have a significant O-SP or capsule component are either non- or weakly pathogenic.[8] Therefore, there is much interest in understanding the role that carbohydrates play in biochemical systems and, particularly, the immune system.

1.2 Glycobiology and immunology

The primary purpose of the immune system of an organism is to detect a wide range of substances and to differentiate between those that are potentially dangerous and those that are not. In this respect, pathogenic bacteria are identified based on specific structures (antigens) found on their cellular surfaces, allowing for their targeted destruction. Cell surface glycosylation is a characteristic of almost all cells and, as a result, cell surface polysaccharide play a key role in the process of antigen recognition. This is as the location of cell surface polysaccharides at the outermost portion of the bacterial cell surface readily allows for their interaction with the host's immune system. The topics of immunology and glycobiology are very broad and only a brief overview will be provided in this Section.

Cell surface polysaccharides are a major determinant of virulence for a large number of bacteria. Bacterial cell surface polysaccharide assist in masking cell antigenic surface components, such as cell surface proteins, from the host immune system inhibiting complement activation.[17,19,20] It has also been shown that the net negative charge of CPSs assists in conferring immune resistance and similarly the more negatively charged the capsule is, the greater the ability of the bacterial cell to inhibit phagocytosis. Cell surface polysaccharides, particularly those containing sialic acid,[21] are also capable of mimicking host antigens disrupting the ability of the immune system to differentiate between self and non-self cells. In addition, cell surface polysaccharides can also help prevent opsonization as well as complement-mediated killing if the bacterium is phagocytosed.[3]

To protect the body from invading bacterial pathogens the human immune system utilises three main classes of white blood cells: phagocytes, T-cells, and B-cells. Phagocytes are cells specialised in engulfing foreign substances that are potentially harmful, such as pathogenic bacteria, and destroying them. They may also undertake antigen presentation whereby the antigenic parts of the ingested substance are transferred to the cell surface for display to other cells in the immune system.[22] This presentation is typically done by major histocompatibility complex class II (MHC-II) molecules, which allows for antigen specific $CD4^+$ T-cells to bind to the complex via their T-cell receptor (TCR) and become activated.

T-cells, in general, can be classified into two main groups: Killer T-cells and helper T (T_H) cells. The former of the two is directly able to target and destroy infected cells. T_H cells, on the other hand, play a wider role in immune response. They are capable of detecting antigens displayed on the surface of antigen-presenting cells (APC), such as phagocytes, and then activate and regulate other cells of the immune system to elicit a response. There are also several types of B-cells which may either be activated directly by free floating antigens or by APC.[23] They are involved in antibody production.

The cell surface polysaccharides of a significant number of pathogenic bacteria stimulate B-cells directly without prior activation of T-cells and are thus known as Thymus-Independent (T_I) antigens.[24–29] Cell surface polysaccharides can be divided into two antigenic groups, type 1 antigen (T_I-1) and type 2 antigen (T_I-2) that differ in the manner they activate B-cells. T_I-1 antigens, such as bacterial LPS, can activate B cells in a polyclonal manner. In contrast, T_I-2 antigens, such as bacterial CPS, induce B-cell activation and antibody production and may require some T-cell help.[30] In addition, the mechanisms by which carbohydrate antigens induce an immune response are influenced by other factors including structural conformation, conjugation to proteins as well as the presence and nature of molecular charge.[16,31] Figure 1.3a provides an illustration of T_I-1 B-cell activation and the resulting response by antibody formation.

Polysaccharide molecules are known to bind to B-cells of appropriate specificity by cross-linking 15-20 B-cell surface immunoglobulin (sIg) B-cell receptors (BCR). This binding activates the B-cell via a number of intra-cellular mechanisms, however, it is not yet able to secrete antibodies. Further signalling is required before the B-cell matures to plasma cells and secretes antibodies. The process by which this occurs is still not fully known but it is widely understood to be T-cell independent.[3,34] Although, it has been shown recently that T-cells may play a role in this signalling process and that the nature of the polysaccharide antigen also plays a crucial role in determining the pathway activated.[33,35]

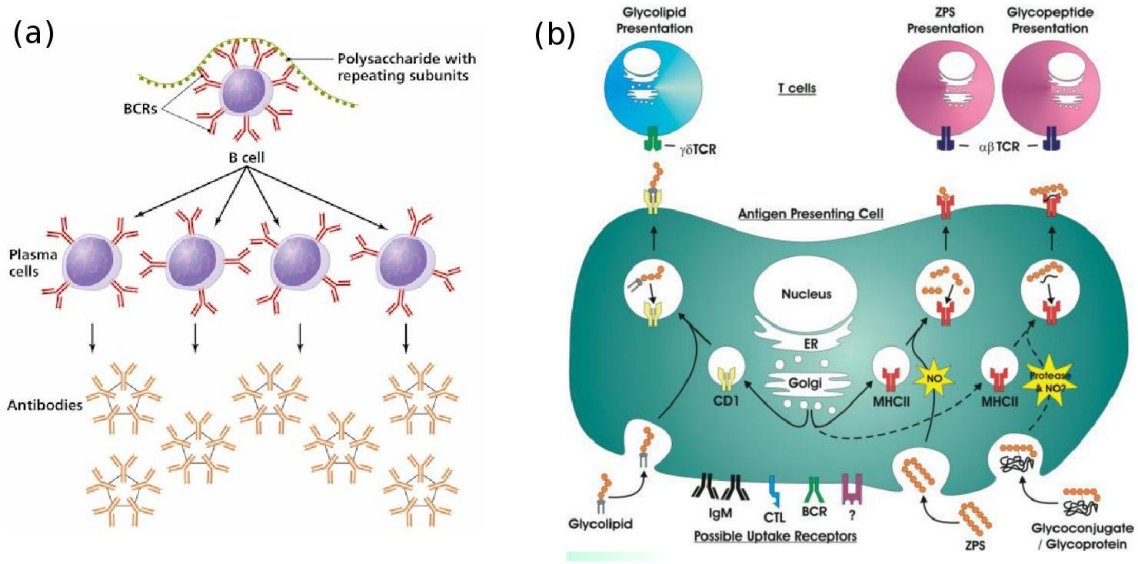


Figure 1.3: (a) T_I-1 B-cell activation and antibody formation by CPS.[32] (b) Three known carbohydrate antigen pathways utilised by the adaptive immune system. Reproduced with permission from Brian A. Cobb, Dennis L. Kasper (2005), © European Journal of Immunology.[33]

For example, polysaccharide antigens that are naturally bound to lipids and peptides, forming glycolipid and glycopeptide complexes, activate the T-cell dependent arm of the immune system.[36] These antigens first bind to sIg receptors on B-cells, are internalised into the cell and the antigen components are then presented on MHC-II complex on the surface. Zwitterionic polysaccharides are also able to activate T-cells although by a different pathway. Three known carbohydrate antigen pathways utilised by the adaptive immune system are illustrated in Figure 1.3b.[33]

1.3 Carbohydrate-based vaccines

The earliest modern record associating cell surface polysaccharide as an immonoreactive antigenic component was in the 1920s by Heidelberger and Avery in their research on *S. pneumoniae*.[37,38] They discovered that a soluble specific substance that was most likely a polysaccharide was located on the cell surface of pneumococci, that it was serotype specific and later found that it could be used as a vaccine providing long lasting immunity. This was supported by further research over the next several years and in 1947 the first licensed carbohydrate-based vaccine was released against six strains of *S. pneumoniae*. However the emergence of antibiotics during the same period, and its preference as a treatment by clinicians, led to a significant decline in vaccine research.[39–42]

Over many subsequent years this slowly began to change, as the efficacy of carbohydrate-based vaccines, particularly toward rapidly growing pathogens, was demonstrated.[39,43,44] The last several decades has also seen a gradual increase in bacterial resistance to antibiotics whilst, in contrast, the number of novel antibiotics and drugs being released to treat infectious diseases has continued to fall. There has thus been a renewed interest in preventative vaccination; of which carbohydrate-based vaccines play an important role. A brief overview of notable historic achievements in the field of carbohydrate-based vaccine research is provided in Figure 1.4.

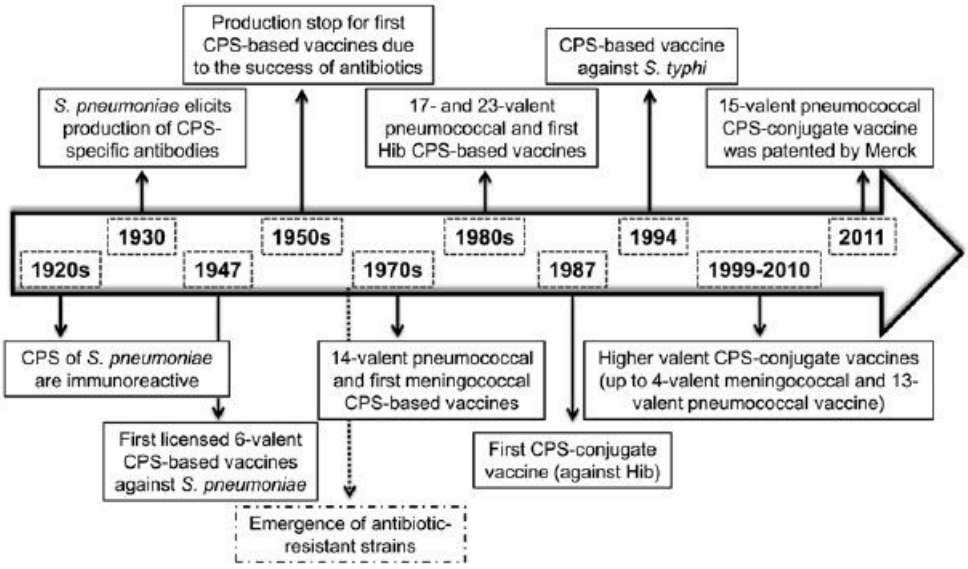


Figure 1.4: Notable history in the development of carbohydrate-based vaccines.[42]

There are two main forms of carbohydrate-based vaccines: polysaccharide vaccines and glycoconjugate (carbohydrate-protein conjugate) vaccines. Polysaccharide vaccines are a unique type of inactivated subunit vaccine that utilises bacterial cell surface polysaccharide as the primary antigenic component. Glycoconjugate vaccines, on the other hand, consist of a polysaccharide or oligosaccharide component linked to a suitably antigenic protein carrier. Polysaccharides and conjugate vaccines activate separate arms of the immune system providing different levels of protection. Table 1.2 details the incident rate and the availability of carbohydrate-based vaccines for a selected number of clinically significant bacterial pathogens.

The key antigenic component of most of the pathogens listed in Table 1.2 are bacterial cell surface

Table 1.2: Estimated annual number of cases and death from selected pathogens.[3, a] Death rate and vaccines statistics updated with WHO and FDA data.

Organism (bacterial)	Disease	Death rate or morbidity	Vaccine
Streptococcus pneumoniae	Acute respiratory infections and meningitis	1-2 million deaths	Polysaccharide and conjugate vaccines available.[45,46]
Haemophilus influenzae	Acute respiratory infections and neonatal meningitis	300 000 to 500 000 deaths	Conjugate vaccines available.[47,48]
Neisseria meningitidis	Meningitis and bacteraemia	Over 500 000 cases, 50 000 deaths (additional 60 000 left with neurological damage)	Polysaccharide and conjugate vaccines available.
Shigella	Shigellosis – diarrhoea and dysentery	Over 100 million cases, 100 000 deaths	Conjugate vaccines available for certain serogroups[49,50]
Salmonella enterica ssp *typhi*	Typhoid (all ages)	21.6 million cases: 216 500 deaths (2004 estimate) 16 million cases: 600 000 deaths (1984 estimate)	Polysaccharide vaccine available: conjugates in development

a for children under five unless specified otherwise

polysaccharides. Consequently, carbohydrate-based vaccines have either been developed or are currently being researched as a means of disease prevention. Both polysaccharide and conjugated vaccines utilise RU fragments from the O-SP or CPS as antigenic epitopes. Table 1.3 provides the O-SP or CPS RU structures of vaccine relevant pathogenic bacteria. It is important to note that an organism may consist of numerous serotypes or serogroups and that the structures listed are a small selection of those that are known to exist.

Table 1.3: Repeating unit structures of selected important bacterial cell surface polysaccharides involved in vaccine development.[3]

Organism	Serogroup/ Serotype	CPS Repeating Unit Structure
Haemophilus influenzae	type b	→3)-β-D-Rib*f*-(1→1)-D- Ribitol-(5→OPO_3→
N. meningitidis	group A	→6-α-D-Man*p*NAc(3/4OAc)-(1→OPO_3→
	group B	→8)-α-D-Neu*p*5Ac-(2→
	group C	→9)-α-D-Neu*p*5Ac(7/8OAc)-(2→
	group W	→6)-α-D-Gal*p*(1→4)-α-D-NeuNAc*p*(7/9OAc)-(2→
	group X	→4)-α-D-Glc*p*NAc-(1→OPO_3→
	group Y	→6)-α-D-Glc*p*(1→4)-α-D-NeuNAc*p*(7/9OAc)-(2→
S. pneumoniae	type 1	→3)-D-AAT-α-Gal*p*-(1→4)-α-D-Gal*p*A(2/3OAc)-(1→3)-α-D-Gal*p*A-(1→
	type 6B	→2)-α-D-Gal*p*-(1→3)-α-D-Glcp-(1→3)-α-D-Rha*p*-(1→4)-D-ribitol-5-P-(O→
	type 7F	→6)-α-D-Gal*p*-(1→2)-[α-D-Gal*p*-(1→3)]-(1→3)-β-L-Rha*p*-[2-OAc] (1→4)-β-D-Glc*p*-(1→3)-β-D-Gal*p*NAc- [-α-D-Gal*p*NAc-(1→2)-α-L-Rha*p*-(1→4)-]-(1→
Shigella boydii O1		→4)-α-D-Glc*p*NAc-(1→3)-α-L-Rha*p*NAc(1→3)-α-D-Glc*p*NAc (1→3)-α-L-Rha*p*NAc(1→3)-β-L-Glc*p*-1-*P*-(O→
Shigella dysenteriae	type O1	→3)-α-L-Rha*p*(1→3)-α-D-Rhap(1→2)-α-D-Gal*p*(1→3)-β-D-GlcNAc*p*-(1→
Shigella flexneri	type 2a	→2)-α-L-Rha*p*(1→2)-α-L-Rha*p*(1→3)-α-L-Rhap(1→3)-β-D-Glc*p*NAc-(1→
Shigella sonnei		→4-α-L-AltNAc-(1→3)-β-D-FucNAc4N(1→

Polysaccharide vaccines

Polysaccharide vaccines are available against *Salmonella typhi*, *Streptococcus pneumoniae* and meningococcal disease although *S. typhi* vaccines are no longer available in many countries. As polysaccharide vaccines are solely carbohydrate based the immune response invoked is most commonly T-cell independent. Upon binding to BCRs polysaccharide vaccine antigens induce B-cells to differentiate into plasma cells and secrete antibodies. However. there is no memory B-cell formation and as a result immunity from polysaccharide antigens are short lived. Figure 1.5 illustrates the B-cell immune response to polysaccharide vaccines.

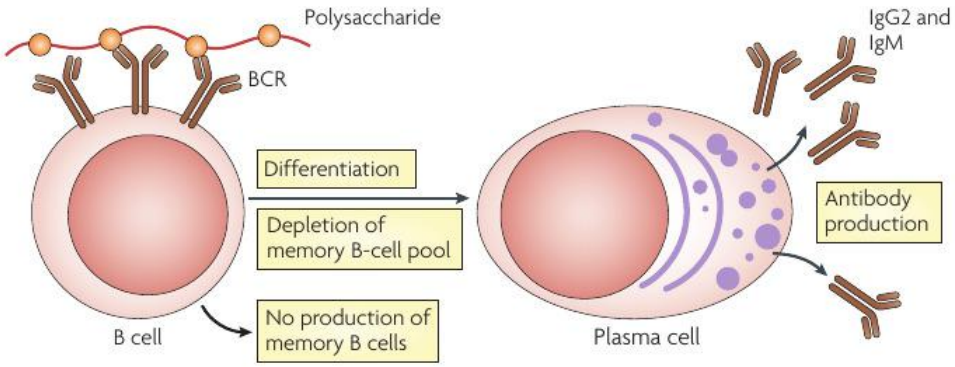


Figure 1.5: B-cell immune response to polysaccharide vaccines.[51]

The ability of a polysaccharide vaccine to stimulate an immune response and produce antibodies is highly dependent on a number of factors. The mass of the polysaccharide has been noted as one factor whereby, as a general rule, polysaccharides with a molar mass of less than 50 000 kDa have been found to be poorly immunogenic whilst those of approximately 90 000 kDa are good immunogens.[52–55] The degree of antibody response to carbohydrate vaccines can also be affected by molecular size, specific determinants as well as the structural conformation of the antigen.[16,56,57]

The age of the individual receiving the vaccine is also an important factor in the efficacy of polysaccharide vaccines. CPS vaccines have been found to be not consistently immunogenic in infants of under 2 years of age as well as in elderly people.[33,58] This is most probably as a result of an immature or compromised immune system.[58] In addition, repeat doses of polysaccharide vaccine in the form of vaccine boosters have not been found to cause a correspondingly significant increase in antibody formation. The use of polysaccharide vaccines thus poses numerous challenges and, as a result, glycoconjugate vaccines have been and are being developed to overcome many of these limitations.

Conjugate Vaccines

Many of the drawbacks of polysaccharide vaccines can be addressed by conjugation of the saccharide to a suitably antigenic protein carrier.[59] This allows for activation of the T-cell dependent arm of the immune system. The activation pathways in response to conjugate vaccines is outlined in Figure 1.6.

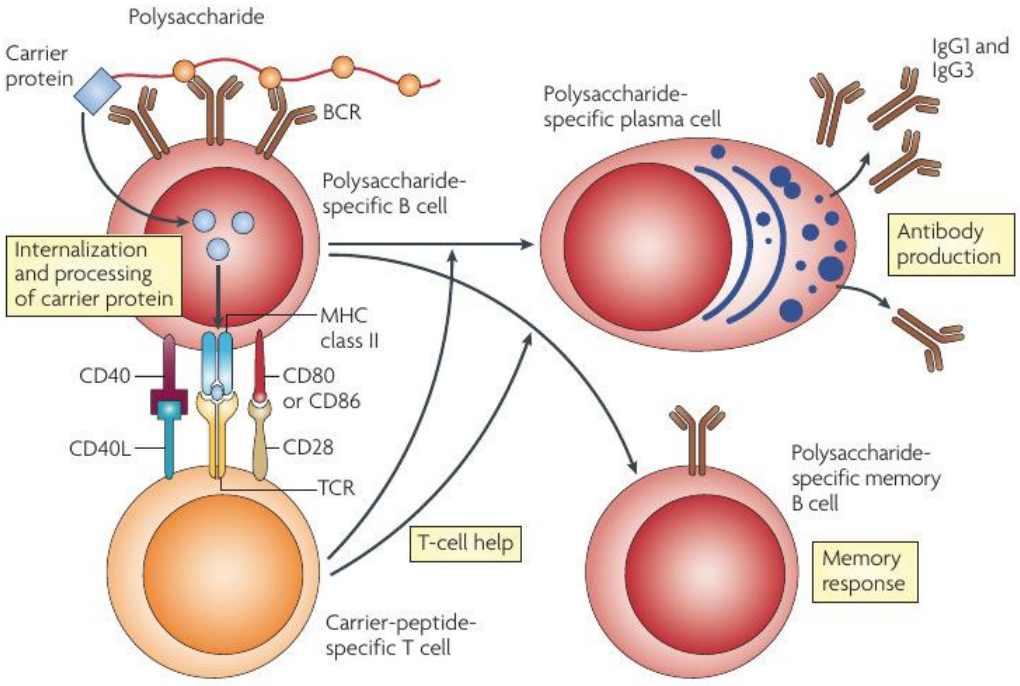


Figure 1.6: Immune response to conjugate vaccines.[51]

Upon recognition by polysaccharide-specific B-cells the carrier protein from conjugate vaccines is internalised and processed . The peptides are subsequently presented on the surface of the B-cell by MHC-II molecules for carrier specific T-cells; resulting in T-cell help for the formation of polysaccharide specific plasma cells and memory B-cells.

The immunogenicity of different conjugate vaccines varies as a result of differences in several factors similar to those of polysaccharide vaccines: the length and conformation of the polysaccharide chain, the charge of the polysaccharide, the amount of unconjugated polysaccharide in the vaccine, and the nature of the carrier protein amongst others.[60,61]

1.4 Meningococcal disease

Meningitis is a disease of the central nervous system caused by inflammation of the meninges - the protective membranes that cover the brain and spinal cord. The disease is primarily caused by bacterial infection of the cerebrospinal fluid, however, it can also develop as a result of viral, fungal, parasitic and, non-infectious causes. Bacterial meningitis is the most acute of these and only pathogenic in humans. It can also lead to severe neurological damage and is fatal in approximately 50% of cases if not treated soon after infection.[62] Three species, *Haemophilus influenzae*, *Streptococcus pneumoniae* and *Neisseria meningitidis* are the leading causative agents of bacterial meningitis.[63,64] These three pathogens are the leading cause of death globally in children infected by vaccine-preventable diseases.[65]

Incidence of the different forms of bacterial meningitis are significantly diverse with infection by *N. meningitidis* (also termed meningococcus) being the most prevalent - especially amongst children and young adults ranging from 1 to 29 years of age.[4,66] Meningococcal disease occurs as a result of infection by meningococcus, a gram-negative β-proteobacterium. The virulence of meningococcus is related to several factors including its genotype and the variability of cell surface structures, especially that of its CPS.[67] The most prevalent groups have demonstrated the ability to readily express and modify their capsules.[67–70] Consequently, meningococcus is unique in its propensity to rapidly spread as wells as cause large scale outbreaks and epidemics. Figure 1.7 provides a cross-sectional view of a meningococcal outer-membrane showing the CPS and other surface associated structures.

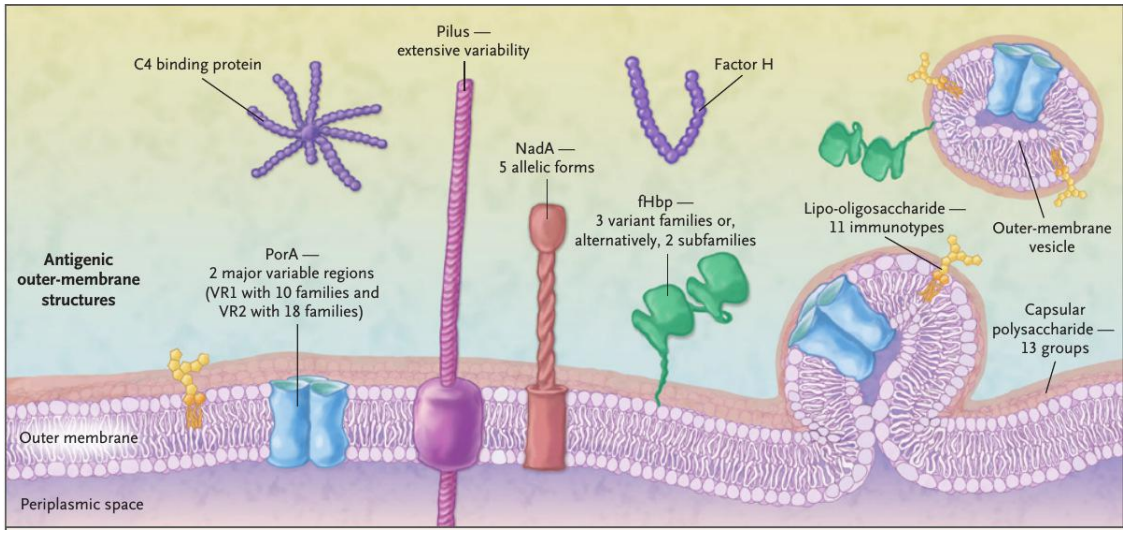


Figure 1.7: Structure of the meningococcal outer-membrane, showing variability of outer-membrane proteins and capsule used in vaccines.[71]

There are 13 known groups of meningococcus based on the immunologic reactivity of their CPSs but only 6 of these (A, B, C, W, X and Y) are the cause of almost all meningococcal disease worldwide. The

meningococcus capsule assists with transmission and colonisation of the disease as well as protecting the bacterium from dessication, opsonisation and phagocytotic and complement-mediated killing.[71–73] In additon, other outer membrane components are also involved in pathogenisis; whereby during invasive disease meningococcus releases LPS and outer-membrane proteins into the bloodstream, potentially diverting the immune system.[74]

1.4.1 Prevalence of meningococcal disease

Meningococcal disease occurs worldwide in both developed and developing countries and is the only form of bacterial meningitis responsible for epidemic levels of infection. *N. meningitidis* can only be carried by and infect humans. The bacteria are carried in the throat, sometimes with no symptoms, and are transmitted from person to person through droplets of respiratory or throat secretions as a result of prolonged close contact. The incubation period is from 2 to 10 days.

The disease can vary in incidence from very rare to over 1000 cases per 100 000 population every year.[67,75–77] It is most prevalent in the regions of Sub-Saharan Africa but incidences elsewhere remain significant even with widespread vaccination regimes.[4,70] However, the degree of incidence and the distribution of the most prevalent groups varies globally. In addition, there has been a recent emergence of additional meningococcal groups in certain regions that have historically had low infection rates. As a result there has been continued study of the mechanisms and epidemiology of meningococcal disease and its causative bacterial agents.

Global prevalence

The most prevalent form of meningococcal disease is caused by group A meningococcus. The incidence of group A disease in developed countries has been low since World War II and has remained so since the introduction of the first meningococcal vaccines in the 1960s.[67] There have been several outbreaks of group A meningococcal (MenA) disease in Russia, China and regions of South East Asia over the past three decades. However, it is most predominant in countries of Saharan and Sub-Saharan Africa where there have been recurring pandemics every 5-10 years since the early 1900s. Although, introduction in 2010 of the first conjugate vaccine against MenA, designed specifically for Africa (MenAfriVac), has led to a significant reduction in meningococcal disease incidence.[62]

Meningococcal groups B (MenB) and C (MenC) are most commonly found in North and South America, Europe and Oceania. They are usually associated with lower incidence rates compared to MenA.[70,77–79] In these regions MenB and MenC are the leading cause of endemic and epidemic meningococcal disease and have caused substantial morbidity and mortality. Figure 1.8 shows the worldwide distribution of meningococcal disease as well as incidence of reported significant outbreaks.

Meningococcal group W (MenW) predominantly occurs in the meningitis belt of Saharan and Sub-Saharan Africa and has also caused multiple meningococcal outbreaks during the Hajj pilgrimage in the 1990s.[64,80] After the outbreak the group W disease spread worldwide and led to a meningococcal meningitis epidemic in Burkina Faso in 2002.[81,82] This was the first recorded large-scale infection of

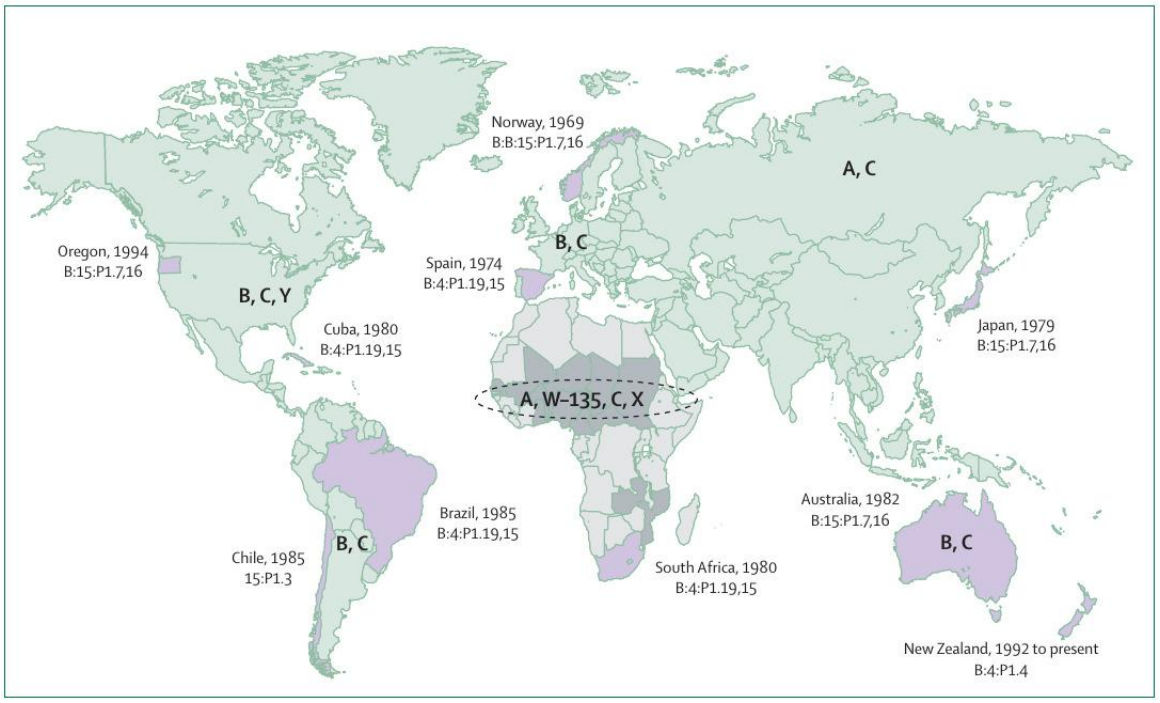


Figure 1.8: Worldwide distribution of major meningococcal groups and recorded outbreaks.[67]

MenW with 13 000 cases reported and 1 400 deaths.[82,83] The past several years has also seen an increase in confirmed cases of MenW disease in several Western European countries.[84,85] England, for example, has recently experienced a consistent increase in invasive meningococcal disease due to MenW bacterium. Cases in the country have increased over five fold year-on-year since 2009, from 22 in 2009 to 117 in 2014, with the incident rate is expected to increase further.[85]

Meningococcal Groups Y (MenY) and X (MenX) have in recent years become more widespread. The incidence rate of MenY has increased since the 1990s in regions such as North America and parts of the Middle East and Asia.[67] MenX has only recently emerged and has caused small outbreaks in Sub-Saharan Africa.[64,67]

Prevalence in Africa

The highest concentration of meningococcal disease worldwide lies in the meningitis belt of Africa (Figure 1.9). The belt can be regarded as a narrow region that crosses the width of the continent comprising Saharan and Sub-Saharan Africa; from Ethiopia in the east to Senegal in the West and reaching as far south as the Democratic Republic of Congo. The epidemiology of the disease in the region is unique with outbreaks occurring every few years in patterns that are atypical with other regions in the world.[67,72]

Five of the six most prevalent groups A, B, C, W and X are found across the meningitis belt with MenA being the most prevalent.[62] In these areas during the dry season, from December to June, populations are at high risk of outbreaks of the disease. The reasons for this trend are still not fully understood, although studies have confirmed that environmental factors such as absolute humidity and dust concentration are

important.[86] MenA remains the most prevalent, however, the incidence rates are falling as a result of widespread use of MenAfriVac over the last several years.[62] Nonetheless, meningitis epidemics due to other groups remain but at lower frequencies and magnitudes.

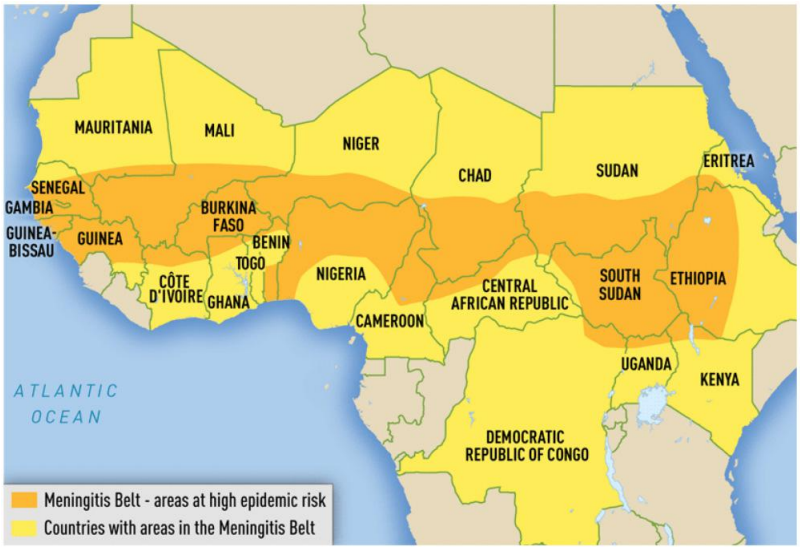


Figure 1.9: Distribution of major meningococcal groups in Africa

1.4.2 Capsular polysaccharides of meningococcus

There are 13 known groups of meningococcal disease. The CPS RU structures of the six most prevalent groups are depicted in Figure 1.10. Within this group there are pairs of structures that have very similar chemical RUs. MenA and MenX CPS RUs are both monosaccharides and they only differ in terms of the type of residue (ManNAc or GlcNAc respectively) and the position of the phosphodiester linkage. MenB and MenC CPS RUs both consist of a sialic acid homopolymer with the major difference being the linkage positions - C-8 for MenB and C-9 for MenC. MenY and MenW both form disaccharide repeating units that contain a sialic acid residue and glucose or galactose respectively. As such, the orientation of the hydroxyl at C-4 is the single variation between the two CPS RUs.

1.4.3 Meningococcal vaccines

The first successful vaccine against meningococcal disease, developed in the early 1960s, was purely polysaccharide based and targeted groups A and C.[76] This was followed a number of years later by protection against additional groups Y and W, which were introduced as a tetravalent A, C, Y, W polysaccharide vaccine[87,88] As results of its safety and immunogenicity a tetravalent polysaccharide vaccine against these four groups remains in wide use today. In addition, there are also mono- and bivalent polysaccharides vaccines employed in regions where they are most in need. One example of this is MenAfriVac against group A meningococcal disease.

However, while polysaccharide meningococcal vaccines have been highly effective in older children and adults with over 85% efficacy, they are generally poorly immunogenic in young children under 2 years of age.[88] Furthermore, meningococcal polysaccharides are T-cell independent antigens that are only able

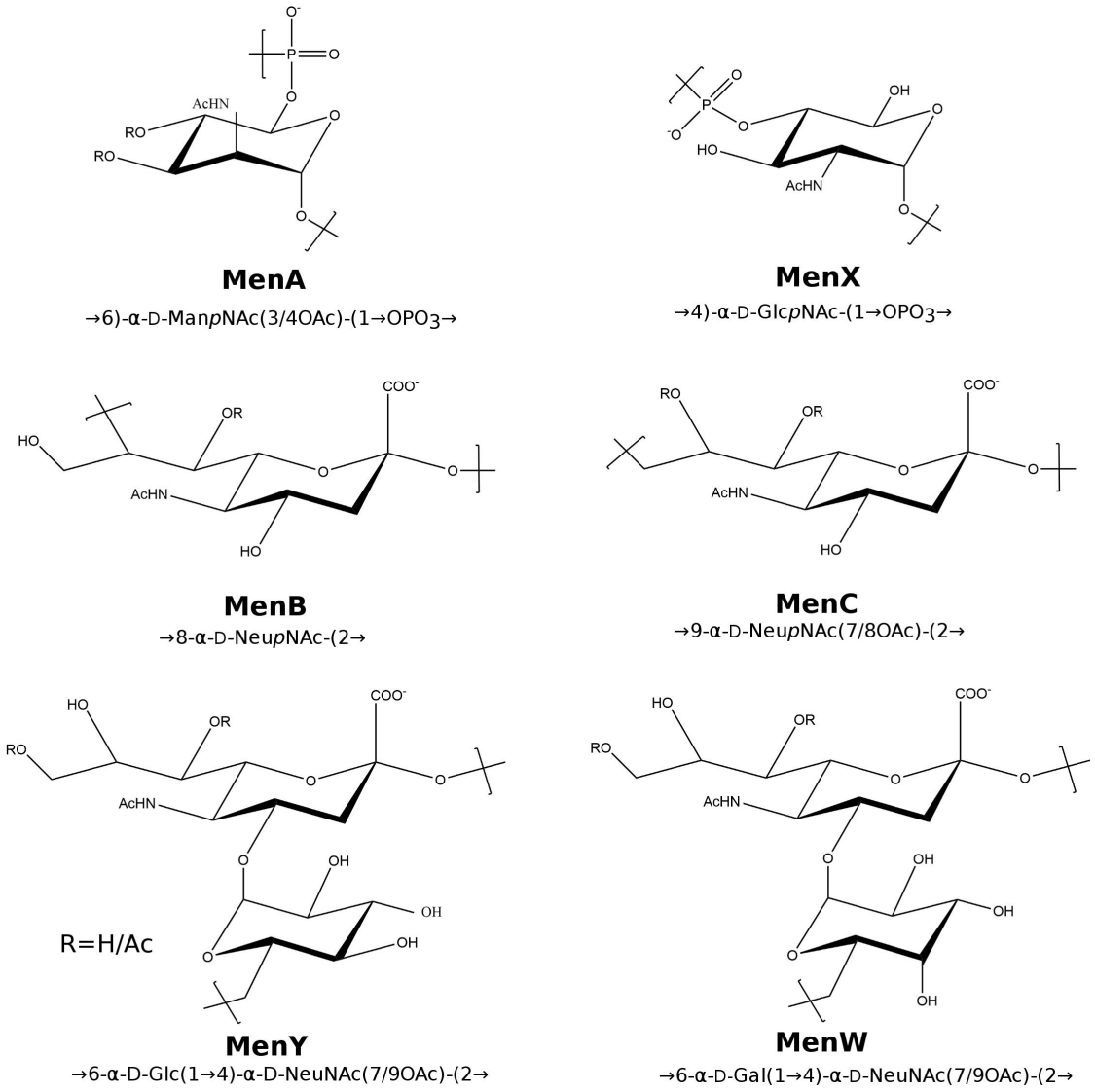


Figure 1.10: Capsular polysaccharide repeating unit structures of the six most prevalent meningococcal groups.

to weakly activate memory B-cells and thus these vaccines induce a relatively short-lived immunity.[51] To overcome these challenges, conjugate vaccines were developed; first against group C then A, Y, and W as well. Both the polysaccharide and conjugate vaccine forms against meningococcal disease are currently in use.

The use meningococcal vaccines over the last few decades have significantly reduced the infection and mortality rates worldwide amongst both adults and children. However, potential challenges have been identified with regards to the continued use of group-specific meningococcal vaccines. One such example provided by Rosenstein et al. suggests that capsule switching may become an important mechanism of virulence due to wide-spread use of vaccines that provide group-specific protection.[4]

1.4.4 Polysaccharide vaccine trials of MenY and MenW

As part of the initial studies exploring possible vaccines against meningococcal disease a small scale clinical trial testing polysaccharide vaccines against meningococcus groups Y and W was conducted.[89] Monovalent and divalent vaccine forms were tested on adult volunteers and the reactogenicity, binding, and bactericidal antibody formation of the vaccines were monitored. The immunological response of volunteers vaccinated with both the monovalent and divalent CPS vaccines as measured by a bactericidal antibody assay are provided in Table 1.4.

Table 1.4: Immunological responses (bactericidal antibody) of volunteers vaccinated with group Y, group W, or a combined Y-W CPS vaccine.[89]

Meningococcal Antigen	Week	% Volunteers responding to meningococcal CPS vaccine		
		MenY Vaccine	MenW Vaccine	MenY-W Vaccine
Y	4	100	30	100
	26	89	25	90
W	4	71	90	80
	26	57	80	80

The results show that both the MenY and MenW monovalent CPS vaccines were able to elicit a strong immune response against their respective antigens. All the volunteers tested with the Y CPS vaccine elicited a significant immune response toward the Y bacterial antigen. Similarly, a very good immune response was induced by the W CPS vaccine toward the W bacterial antigen; with 90% of volunteers responding when measured for bactericidal antibody after 4 weeks. The divalent vaccine, as expected, was able to protect well against both bacterial antigens.

However, when testing the vaccines for cross-protection against the opposing meningococcal antigen a notable result was found. The MenY CPS vaccine elicited a good immune response against the group W bacterial antigen with 71% of volunteers responding after 4 weeks. In contrast, the MenW CPS vaccine had a comparatively weak action against the MenY antigen with 30% of volunteers responding after the same period. This result was not readily explained as the primary repeating unit structures of the CPS RUs for the two groups are very similar.

A more recent study provides further support for the possible cross-protection by the group Y vaccine. A clinical trial of a HibMenCY-TT conjugate vaccine showed markedly higher seroprotection rates and antibody titre to group W compared to the control group, even though neither group had previously received the group W antigen. The lack of response in controls indicates that the high group W responses observed in the HibMenCY-TT group are not the result of natural immunity or a lack of assay sensitivity. The results from these two trials are the motivation for the computational component in this study, whereby the CPS RU of MenY and MenW are modelled to determine if there is a correlation between conformation and CPS vaccine activity.

1.5 Shigellosis

Approximately one in ten deaths of children under five worldwide is due to diarrhoeal or dysentery related diseases,† resulting in almost 800 000 fatalities annually.[90] A global multi-year study ending in 2013 found that these incidences were primarily due to infections from four pathogens: rotavirus, *Cryptosporidium*, enterotoxigenic *Escherichia coli*, and *Shigella*.[90] These diseases can be endemic or epidemic and are significantly affected by levels of hygiene, sanitation, and access to healthcare. As such, infections are predominant in the developing world, however, with continued improvement in education, healthcare, and socio-economic development incidences have steadily decreased over the last decade.

Nonetheless, mortality rates continue to remain high and as a result prevention and treatment of these disease remains universally important. In this respect, there are several effective means for preventing and treating dysentery related diseases. The development of effective multivalent vaccines forms part of this strategy. Vaccines have been developed for rotavirus and *Escherichia coli* and as of 2015 multiple multivalent vaccine against shigellosis (*Shigella* disease) are in development. For children of 1-2 years and 2-5 years *Shigella* ranks amongst the leading attributable incidences of dysentery (Figure 1.11).

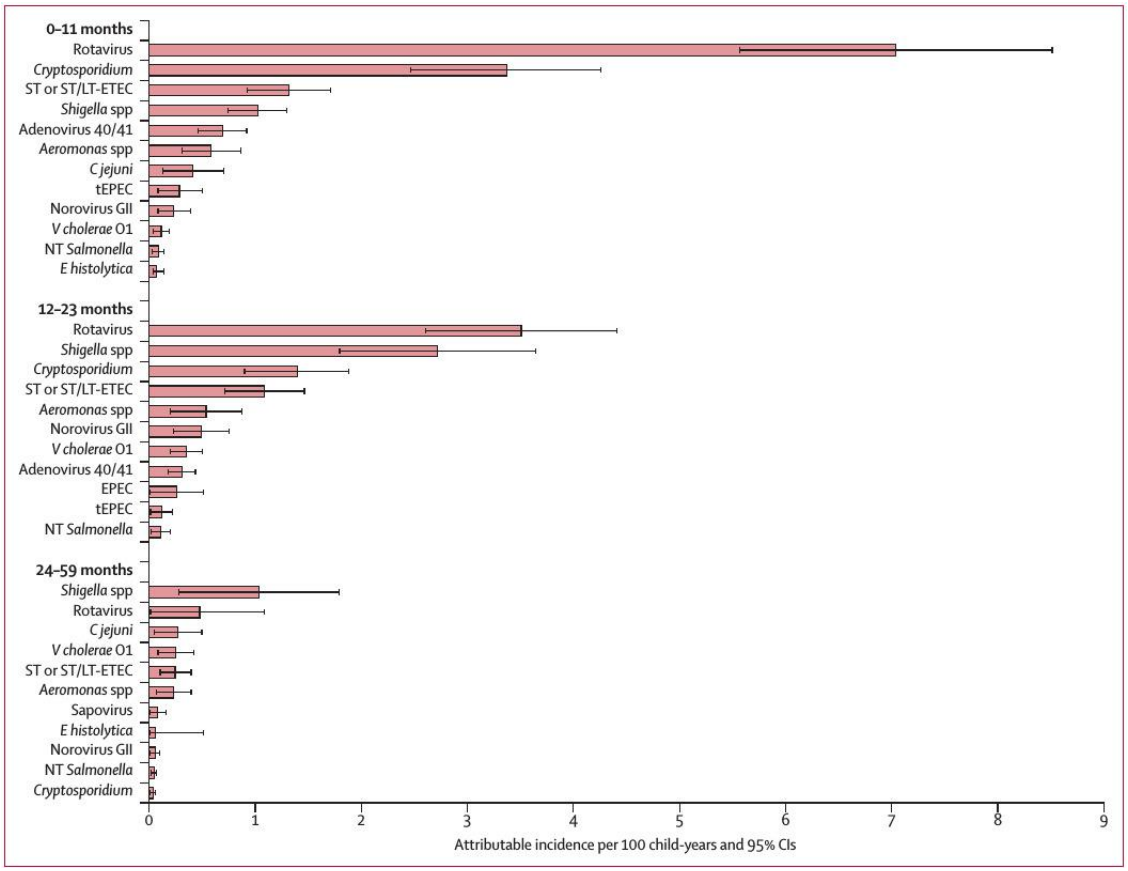


Figure 1.11: Attributable incidence of pathogen-specific dysentery per 100 child-years by age. The bars show the incidence rates with the 95% confidence interval indicated by the error bars.[90]

† Diarrhoea is defined as four or more watery bowel movements daily whilst dysentery is defined as fever, cramps, and the presence of blood alone or with mucus, in stool. Dysentery is generally associated with higher mortality rates and has a greater effect on long term growth and retardation than diarrhoea.

First discovered in 1897, *Shigella* is a genus of gram-negative, non-motile, rod-shaped, enteric bacteria that typically causes dysentery upon infection in humans.[91,92] There is an estimated 80-165 million cases of infection by *Shigella* annually, with at least 100,000 of these resulting in death each year.[93] The *Shigella* species is classified into four groups based on common O-SP RU structures and, with the exception of *S. sonnei*, each contain multiple serotypes:

Group A - *S. dysenteriae*: 15 serotypes
Group B - *S. flexneri*: 6 serotypes
Group C - *S. boydii*: 19 serotypes
Group D - *S. sonnei*: 1 serotype

There are a number of methods for reducing fatalities from dysentery related diseases. Economic advancement and education and the corresponding improvement in hygiene and sanitation arguably play the greatest role, however, there remains a strong need for effective drugs and vaccines.[94,95] *Shigella* bacteria are one of a few genus of pathogens that have had their pathogenesis and biochemical interactions with mammalian cells studied extensively at both the cellular and subcellular levels.[96–98] Nevertheless, progress in attaining a safe and effective vaccine against the most prevalent serotypes of *Shigella* has proved challenging.[95]

1.5.1 Prevalence of shigellosis

Global Prevalence

Shigellosis is primarily spread via contaminated food or water and thrives in regions with high population densities and inadequate sanitation. As such, shigellosis predominantly occurs in the developing world where it is concentrated in regions that allow *Shigella* bacterium to thrive. A recent study by the World Health Organisation (WHO) found that as much as ninety-nine percent of all incidences of shigellosis occurs in the less developed countries and infection is primarily in children under 5 years of age.[93] In addition, these infections were found to be more severe and have higher cases of mortality. The regions where shigellosis is most prevalent include South America, the Sub-Continent and parts of North and Sub-Saharan Africa. Figure 1.12 illustrates the global distribution of the predominant shigella species.

Globally *S. flexneri* is the predominant cause of shigellosis while the strain that results deadly epidemics in the developing world is *S. dysenteriae*. In more industrialised regions, such as those of North America, Europe and Oceania, *S. sonnei* is the prominent serotypic cause of shigellosis.[95]. For example, over two thirds of the approximately 500 000 incidences of shigellosis in the United States is due to *S. sonnei*.[100] In addition, *S. sonnei* has been strongly linked as being a leading cause of traveller's diarrhoea.[101]

Prevalence in Africa

In Africa shigellosis has the highest incidence in the east and western regions of Sub-Saharan Africa. These include countries such as Somalia, Ethiopia, Kenya, and Tanzania in the East and Ghana, Mali, Cameroon, and Nigeria in the West. It is less commonly found in the southern regions of Africa.

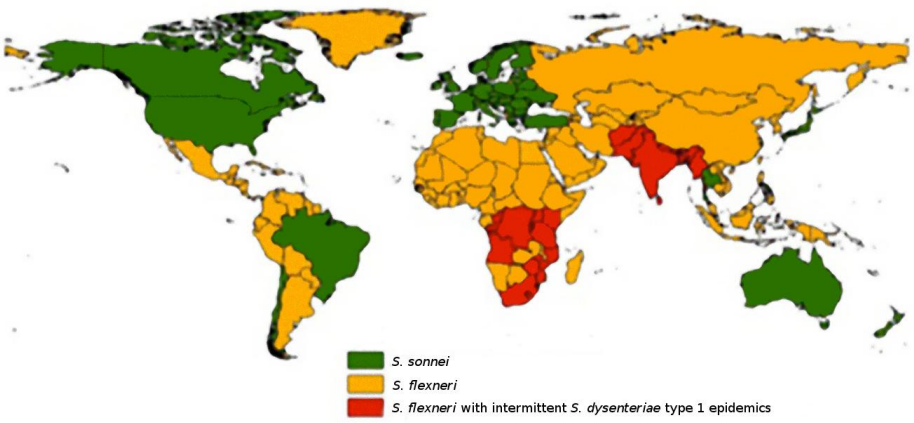


Figure 1.12: Global distribution of predominant shigella species.[99]

Nonetheless, as shigellosis is a food and water borne disease, it particularly thrives in regions of overcrowding, poor sanitation and a lower quality of healthcare. As a result of these epidemiological niches, incident rates remain comparatively higher in Africa than more developed regions.

1.5.2 O-specific polysaccharides of *Shigella*

There are four known groups of *Shigella*. The O-SP RU structures of the bacterium varies significantly between groups. The RU structure of *S boydii* O1 is a pentsaccharide consisting of glucose and rhamnose derivatives. *S. dysenteriae* O1 and *S. flexneri* 2a contain tetrasaccharide RUs with N-acetylglucose, galactose and rhamnose. *S. sonnei* is the only serotype that contains a disaccharide O-SP RU structure. Table 1.5 lists the CPS RU structures of the four *Shigella* groups.

Table 1.5: Repeating unit structures of the four major serotypes of *Shigella*.[102–105]

Serotype	O-SP Repeating Unit Structure
Shigella boydii O1	→4-α-D-Glc*p*OAc-(1→2)-α-L-Rha*p*OAc(1→3)-β-L-Rha*p*(1→4) -β-L-Rha*p*(1→3)-β-D-Glc*p*-1-*P*-(O→
Shigella dysenteriae O1	→3-α-L-Rha*p*(1→3)-α-D-Rha*p*(1→2)-α-D-Gal*p*(1→3)-β-D-GlcNAc-(1→
Shigella flexneri 2a	→2-α-L-Rha*p*(1→2)-α-L-Rha*p*(1→3)-α-L-Rha*p*(1→3)-β-D-Glc*p*NAc-(1→
Shigella sonnei	→4-α-L-AltNAcA-(1→3)-β-D-FucNAc4N(1→

1.5.3 Vaccines against *Shigella*

There are currently no licensed vaccines to protect against shigellosis. Over the last several years there has been an attempt to develop a multivalent conjugate vaccine against *Shigella* by providing protection against the most prevalent serotypes: *S. dysenteriae* O1, *S. flexneri* 2a, *S. flexneri* 3a, *S. flexneri* 6, and *S. sonnei*.[95] A key virulence factor of the *Shigella* bacterium is the organism's O-SP, and it is known to elicit a strong immune response in humans.[106] The O-SP has thus been used as an antigenic component in the development of carbohydrate based conjugate vaccines.

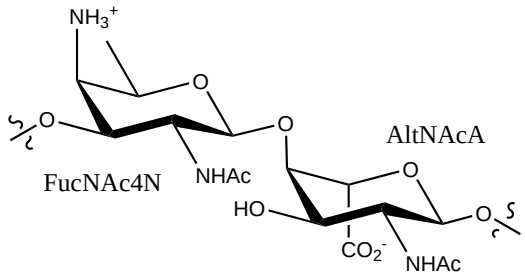


Figure 1.13: O-Specific polysaccharide repeating unit of *S. sonnei*.

There are numerous barriers to the production of conjugate vaccines with the process often being complex and expensive.[44] The antigenic polysaccharides and protein carriers are biosynthesized in bacteria, biologically or chemically cross-linked, and purified in multiple steps. Development of conjugate vaccines also requires biological testing and physicochemical characterisation of the manufactured products. This is typically done against synthetic samples and, as such, requires a sufficient quantity of pure synthetic carbohydrate antigen. Therefore, the synthesis of the O-SP RUs in a cost effective manner and in a sufficiently large scale is of importance.

Synthesis of the CPS RU of *S. sonnei* poses particular challenges. The zwitterionic polysaccharide has a disaccharide RU made of two uncommon amino- sugars, a 2-acetamido-2-deoxy-L-altruronic acid (AltNAcA) and a 2-acetamido-4-amino-2,4,6-trideoxy-D-galactopyranose (FucNAc4N) 1,2-trans linked to each (Figure 1.13).[104] The *S. sonnei* O-SP disaccharide RU also consists of three amino groups, one of which is zwitterionic and the remaining two occur as acetamides. This chemical characteristic is not commonly found in bacterial polysaccharides and thus makes their synthesis of particular interest.

1.6 Scope of work

This study has two main parts. The first comprises computational modelling of meningococcal groups Y and W CPS RUs. The second is an experimental component, which is part of a broader research project to develop a novel synthetic route towards the O-SP RU of *S. sonnei*.

Aims

The aim of the molecular modelling part of the study is to determine if the different levels of antibody cross-protection observed in the immune response to meningococcal Y and W polysaccharide vaccines can be attributed to CPS RU conformation. The aim of the experimental section is to synthesise the protected form of the FucNAc4N residue present in the O-SP RU of *S. sonnei*.

Objectives

1. Identify preferred orientations of glycosidic linkages in the disaccharide repeating units of MenY and MenW;
2. Perform computational simulations on the RUs and undertake conformational analysis;
3. Attempt to correlate structural observations to the different levels of immunogenicity observed; and
4. Synthesis of the D-FucNAc4N residue of *S. sonnei* from glucosamine hydrochloride starting material.

The structure of the thesis is briefly outlined below:

Chapter 1 is this Introduction;
Chapter 2 outlines the molecular modelling methods used in this study;
Chapter 3 provides background and results for the computational studies performed;
Chapter 4 provides background and results for the synthesis component of the study;
Chapter 5 details the experimental methods for the synthesis performed; and
Chapter 6 gives the conclusions of the study.

Chapter 2. Molecular modelling and analytical methods

2.1 Molecular mechanics

Molecular mechanics (MM) is a process by which systems are studied using classical (Newtonian) mechanics. The potential energy of the system is described by a series of functional forms and parameter sets in what is termed a force field. These parameters that describe the potential energies of different aspects of the system can be derived empirically or from quantum mechanical calculations. The accuracy of the calculations in molecular mechanics simulations is highly dependant on the accuracy of the force field that describes it. In this study we use the carbohydrate optimised CHARMM36 force field.[107–110]

2.2 Force fields

A force field is a set of functions and parameters that describe the potential energy of a particular system of particles. A basic modified functional form of the CHARMM36 force field[111] describing the potential energy, V, of a molecular system with respect to its individual components is given by:

$$V_{pot} = \sum_{Bonds} v_{Stretch} + \sum_{Angles} v_{Bend} + \sum_{Dihedrals} v_{Torsion} + \sum_{Urey-Bradley} v_{1,3-interaction} + \sum_{Pairs} v_{Non-Bonded} \quad (2.1)$$

Described in the individual energy components for the internal (bonding) and external (non-bonding) interactions for atom pair i, j:

Bonding terms

Bonds: $$v_{Stretch} = k_{stretch}(r - r_0)^2 \quad (2.2)$$

where r is the atom pair distance, r_0 is the atom pair equilibrium distance and $k_{stretch}$ is the bond force constant.

Angles: $$v_{Bend} = k_{bend}(\theta - \theta_0)^2 \quad (2.3)$$

where θ is the angle of the three atom group, θ_0 is the equilibrium angle of the three atom group and k_{bend} is the angle force constant.

Dihedrals: $$v_{Torsion} = k_{proper}[1 + cos(n\phi - \delta) + k_{improper}(\omega - \omega_0)^2 \quad (2.4)$$

where ϕ is the dihedral angle, δ phase shift, k_{proper} is the proper dihedral force constant and $k_{improper}$ is the improper dihedral force constant.

1,3 interactions: $$v_{1,3-interaction} = k_{UB}(u - u_0)^2 \quad (2.5)$$

u is the measured 1,3 atom distance, u_0 is the equilibrium 1,3 atom distance in the harmonic potential and k_{UB} is the Urey-Bradley force constant.

Non-Bonding term

$$Nonbonded: \qquad v_{Non-Bonded} = \varepsilon_0 \left[\left(\frac{R_{ij}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j e^2}{r_{ij}} \tag{2.6}$$

The non-bonded term is composed of the van der Waals energy, as described by the 12-6 Leonard-Jones potential, and the standard Coulombic potential. R is the distance at which the 12-6 Leonard-Jones potential is zero.

2.3 Molecular dynamics

Molecular dynamics is the process by which the time, t, average motion of particles can be calculated by means of Newton's second law of motion as described by:

$$\frac{d^2\mathbf{r}_i}{dt^2} = \frac{\mathbf{F}_i}{m_i} \tag{2.7}$$

where $\mathbf{r}_i$ is the the position of a particle with mass, m_i, that is acted upon by a force $\mathbf{F}_i$. Numerical integration of the equations of motion in increments of time, δt, allows for the forces on a particle to be calculated at each time step. These forces are the sum of the interactions of all the other particles in the simulated system and, with the position and velocities of a particle at a time, t, the new positions and velocities at time $t + \delta t$ can be calculated. At the beginning of a simulation the kinetic energy of the system is initialized by assigning random velocities from a Maxwell-Boltzmann distribution followed by rapid temperature scaling to the specified starting temperature.

The time step employed in a MD simulation is dependent on the fastest motion of the system and which corresponds to the vibration of bonds to hydrogen. The use of time steps that are greater than the vibrational frequency of these bonds can lead to the atomic positions of the hydrogen atoms being calculated imprecisely from one time step to the next and which may result in inaccuracies in the potential energies generated. Consequently, constraint algorithms, such as LINCS[112] or SHAKE[113] can be employed to fix particular bond lengths, such as those to hydrogen, allowing an increase in time step and a saving in computational time if required.

2.3.1 The integrator

The required initial input conditions for a MD run is the potential interaction as a function of atom positions as well as the positions and velocities of all the atoms in the system. At each time step the forces of every atoms is computed as a function of bonded and non-bonded forces as well as restraints if necessary. The basic form of the velocity-verlet algorithm[114] allows for calculation of the positions, $\mathbf{r}$; velocities, $\mathbf{v}$; and accelerations, $\mathbf{a}$ at time $t + \Delta t$ from the same quantities as time t by:

$$\mathbf{r}(t+\Delta t) = \mathbf{r}(t) + \mathbf{v}(t)\Delta t + (\mathbf{a}/2)t + \Delta t^2 \tag{2.8}$$

$$\mathbf{v}(t+\Delta t/2) = \mathbf{r}(t) + (\mathbf{a}/2)t \tag{2.9}$$

$$\mathbf{a}(t+\Delta t) = -(1/m)\Delta V(r(t+\Delta t)) \tag{2.10}$$

$$\mathbf{v}(t+\Delta t) = \mathbf{v}(t+\Delta t) + (1/2)\mathbf{a}(t+\Delta t)\Delta \tag{2.11}$$

2.3.2 The water model

There are many different potential functions for modelling a water molecule in MD simulations. These can be implicit or explicit solvent models and can range from being simple to quite complex. This study uses the TIP3P water model, which is a rigid, three site model.[115] Rigid models constrain the O – H bond, restricting vibration and rotation and allowing for fast simulations where modelling does not require high accuracy for water molecules. Three site models utilise three interaction points for each of the three atoms in water, with each site acting as a point charge and a Lennard-Jones potential acting on the oxygen atom. In the TIP3P water model, partial positive charges are defined on the hydrogen atoms and a negative charge on the oxygen atom. The intermolecular interaction between two water molecules is computed using a Lennard-Jones type potential by a single interaction point per molecule on the oxygen atom.

2.3.3 Periodic boundary conditions

MD simulations often model a small sample of much larger chemical systems. Consequently, the surface area to volume ratio is much greater for the MD sample and, therefore, the particles spend a much greater fraction of time close to the edge of the system than would otherwise normally occur. This results in what is termed 'edge effects' whereby the interactions at the outer portions of the system have a disproportional effect on the overall simulation.

To minimise this effect periodic boundary conditions (PBC) are employed by modelling the simulated system as a unit cell. The particles in a single unit cell is simulated and the motion of particles in cells adjacent to the primary cell replicate the same movement. Therefore, as one particle moves out of the unit due to its simulation motion the same particle moves into the unit cell at the exact opposite point. Effectively this results in the system of the primary unit cell having no walls, significantly reducing the errors that may arise as a result of 'edge effects'. Figure 2.1 illustrates a molecular dynamic system when modelled with and without periodic boundary conditions.

2.4 Metadynamics

Metadynamics simulations can be regarded as a subset of MD simulations and employs many of the same methods. However, one key contrast is the manner by which the two methods explore different

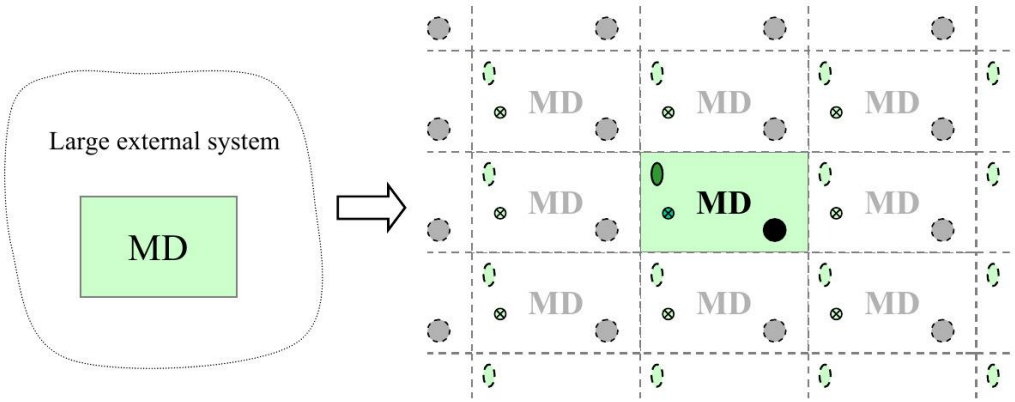


Figure 2.1: MD simulations with (right) and without (left) periodic boundary conditions.[116]

system states. In the case of simulating a molecule using these methods, the system state may be defined with respect to the overall conformation of the molecule.

MD simulates the natural forces within and exhibited on a molecule using classical mechanics and can thus be regarded as a deterministic method. Systems modelled with MD will thus have a natural tendency to move toward lower energy equilibrium conformational states. However, the different states that can be accessed by a molecule over any period of time are entirely dependant on natural variations of system forces. If a molecular conformation resides far away energetically from lower energy equilibrium states, longer periods of simulation time are required or the conformational space may be inadequately sampled.

In contrast, metadynamics employs the application of an external force to certain parts of a system that allows for an enhanced sampling rate of conformational space. The external force can be applied to a certain well defined segment of a system known as a collective variable (CV), such as inter-atomic distances, bond angles, and dihedral angles. During the simulation the potential of the system as a function of of one or more CV, V_{meta}, is measured and a positive Gaussian energy potential is added, reducing the probability of the system returning to its previous state. If the simulation is performed for a sufficient period of time, a proper statistical average over all configurations of the CV is formed. As the free energy measured during this process is a function of one or more CV, it is referred to as a potential of mean force.

The metadynamics potential[117] acting on the collective variable ξ as used in this study is given by:

$$V_{meta}(\xi) = \sum_{t'=\delta t,2\delta t,\ldots}^{t^i<t} W \prod_{i=1}^{N_{cv}} exp\left(\frac{(\xi_i - \xi_i(t'))^2}{2\delta_{\xi_i}^2}\right) \tag{2.12}$$

where V_{meta} is the history-dependent potential at time, t; N is the number of CV; ξ is the value of the CV at t; and W is the Gaussian hill-weight. In this way the potential energy landscape of the CV can be

measured and as this is fully explored $V(cv,t)$ approaches the potential of mean force, $E(\xi)$:

$$E(\xi) \approx \lim_{t\to\infty} V_{meta}(\xi) \tag{2.13}$$

2.5 File formats

This study uses the Protein Data Bank (PDB) and and the CHARMM Protein Structure File (PSF) formats for describing the chemical structures simulated. The PDB format used provides the three dimensional coordinates of all the atoms in the modelled chemical system. The PSF format used details the atom coordinates, the bonding connectivity between the atoms, three-atom angles, four-atom proper and improper dihedral angles as well as cross-terms. These files provide a detailed description of the system simulated and are required for executing the MD simulation as performed in this study.

2.6 Analytical methods

2.6.1 Root mean square deviation

The Root Mean Square Deviation (RMSD) of a chemical structure can be used as a measure of conformational difference relative to a reference structure. Analysis of the change in RMSD of a molecular structure over the course of a MD simulation allows for studying conformational change with time. It is commonly calculated with respect to a simulation starting structure post equilibration of the system and is typically done on a specific groups of atoms. Small variations in the RMSD over time indicates greater conformational stability whilst larger fluctuations suggests a more flexible structure. Convergence of the RMSD toward a particular range during the simulation indicates a molecular structure that is conserved and that has likely settled in a lower energy equilibrium conformation. The weighted by mass RMSD of a structure at a time t_2 relative to a reference structure at time t_1 as used in this study is given by:

$$RMSD(t_1,t_2) = \frac{1}{M}\left[\sum_{i=1}^{N} m_i ||\mathbf{r}_i(t_1) - \mathbf{r}_i(t_2)||^2\right]^{\frac{1}{2}} \tag{2.14}$$

where M is the molar mass, m is the atomic mass, and $\mathbf{r}(t)$ is the position of atom i at time, t.

2.6.2 Radius of gyration

Radius of gyration, R_g, provides a measure of the overall spread of a conformational structure. The smaller the R_g value the more compact the molecule while the larger the value the more indicates it is more extended. The mass weighted R_g as used in this study given by:

$$R_g = \left[\left(\sum_{i=1}^{n} m_i ||\mathbf{r}_i - \mathbf{r}_c||^2\right) / \left(\sum_{i=1}^{n} m_i\right)\right]^{\frac{1}{2}} \tag{2.15}$$

where $\mathbf{r}_i$ and m_i is the position and mass of the ith atom, and $\mathbf{r}_c$ is the position of the centre of mass of the molecule.

2.6.3 Distance

End-to-end distance measurements may be used as an additional method for determining the degree of folding for different molecular conformations. The end-to-end distance, d, used were based on measurements from the ring oxygen atoms of the terminal residues and calculated by:

$$d = \left[(\mathbf{r}_i - \mathbf{r}_j)^2 \right]^{\frac{1}{2}} \tag{2.16}$$

where $\mathbf{r}_i$ and r_j are the positions atom i and atom j.

2.6.4 Cluster analysis

MD simulations generate a configuration of a modelled system at each time step. Typically this results in a large number of configurations, although, many of these are very similar. Conformational analysis of a molecule from a dynamics run requires studying the structure over the course of the simulated trajectory. In this regard, clustering provides a method for grouping of these structures (at different time steps) based on one or more predefined, measurable criteria. Clustering allows for identifying families of structures with similar conformations

There are a number of clustering algorithms that can be used for the conformational analysis of molecules. In this study we use a method based on the quality threshold algorithm.[118] The clusters are calculated by specifying the measurement criteria, the maximum threshold distance as well as the approximate number of clusters to be calculated. Initially, a random point is selected from the sample and a candidate cluster is built around it by iteratively including the next closest point until the threshold distance is surpassed. The cluster with the most points is the first true cluster and these points are removed from the sample for further consideration. This process is repeated with reduced set points until all points are allocated to clusters. If a point cannot fit in any of the formed clusters it is added to an additional outlying group.

2.6.5 Solvent accessible surface area

The solvent-accessible surface area (SASA) is the surface area of a chemical structure that is accessible to a solvent. The parameter was measured as carbohydrate properties, such as glycosidic linkage rotamer conformation, molecular stretching, and ring puckering, are affected by interactions between the carbohydrate and solvent molecules.[119–121] Thus, measuring solvent-accessible area, as well as studying the relationship between changes in molecular conformation and solvent-accessible area, may assist in better understanding carbohydrate structural behaviour.

The SASA calculation is typically performed on each atom (or specific group of atoms) of the molecule

by tracing a sphere with a specified radius around the atom from it's centre. A measurement is then made to find the points on the sphere that are occluded or exposed to solvent, giving the solvent-accessible area around the atom. The area around each atom of the molecule calculated as being solvent exposed are added together to provide the total solvent-accessible surface area of the molecule.

Chapter 3. Conformations of meningococcal Y and W capsular polysaccharide vaccines

3.1 Introduction

MenY [→6)-α-D-Glc(1→4)-α-D-NeuNAc-(2→] and MenW [→6)-α-D-Gal(1→4)-α-D-NeuNAc-(2→] bacteria have very similar CPS RUs with the only difference being the orientation of the C-4 hydroxyl: equatorial for MenY and axial for MenW. Figure 3.1 illustrates de-O-acetylated (dOA) dimers of MenY and MenW CPS as used in this study. It should be noted that the sialic acid residues of the RUs are partially OAc at C-7 and C-9 depending on the strain.[122] Only a relatively small proportion of the MenY and MenW CPS RUs are OAc (approximately 15%[122]) and therefore we do not consider the effects of OAc in this investigation. This approach was also taken in the previous modelling study performed on MenY and MenW.[57]

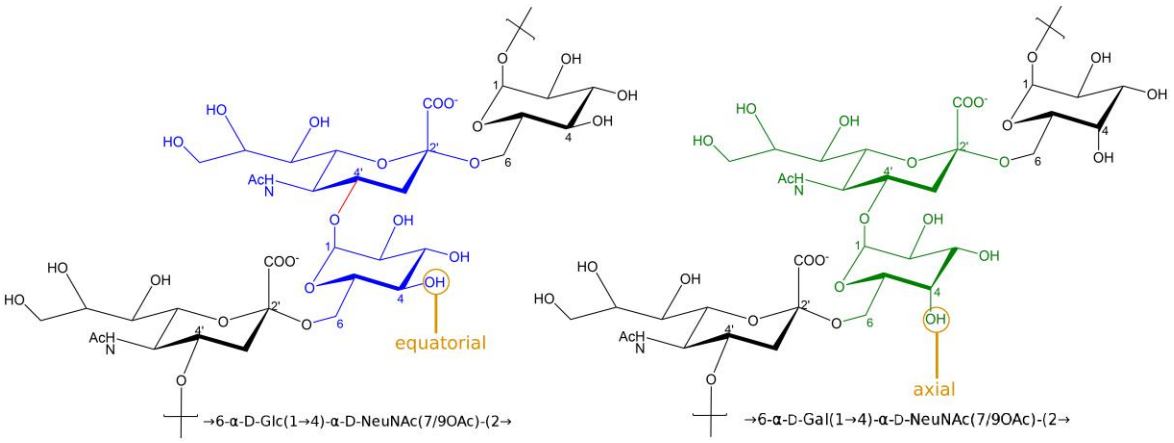


Figure 3.1: De-O-acetylated CPS dimer of MenY (RU left in blue) and MenW (RU right in green).

A small scale clinical trial by Griffis *et al*. showed that MenY and MenW CPS vaccines provide immune cross-protection against bacteria of the other meningococcal group (Table 3.1).[89]

Table 3.1: Immunogenicity of meningococcal group Y and W CPS vaccines.[89]

Antigen	Y Vaccine	W Vaccine	Y-W Vaccine
Y	Very Strong	Weak	Very Strong
W	Strong	Very Strong	Very Strong

The study showed that the MenY CPS vaccine was able to provide high levels of bactericidal antibody against group W bacteria but that the MenW CPS vaccine was only able to weakly cross-protect against group Y bacteria. The reasons for the observed difference are not fully understood, however, the relationship between capsular polysaccharide conformation and immunogenicity has been previously shown.[16,123]

Moore *et al*. modelled the MenY and MenW CPSs as part of a study investigating the nature of group

specific epitopes.[57] Their simulations employed molecular mechanics (utilising the GEometry of GlycOProteins package and a modified Hard Sphere Exo-Anomeric force field)[124] in combination with a Metropolis Monte-Carlo protocol to investigate molecular conformation. The global minimum of each individual glycosidic linkage was calculated in vacuum to determine their preferred low energy orientations. These were then used to build larger conformations, which were further minimised. The process was repeated up to a length of 10 RUs.

The modelling predicted similar helical structures for Men Y and MenW, with four residues per turn, and the largest difference being the orientation of the C-4 hydroxyl. In the MenW 3RU the C-4 hydroxyl is positioned such that it lies in a hydrophobic surface region. In the MenY RU the C-4 hydroxyl disrupts this surface, increasing the hydrophilic nature of the molecule in comparison to that of MenW. Figure 3.2 provides a model of the simulated MenY and MenW RUs.

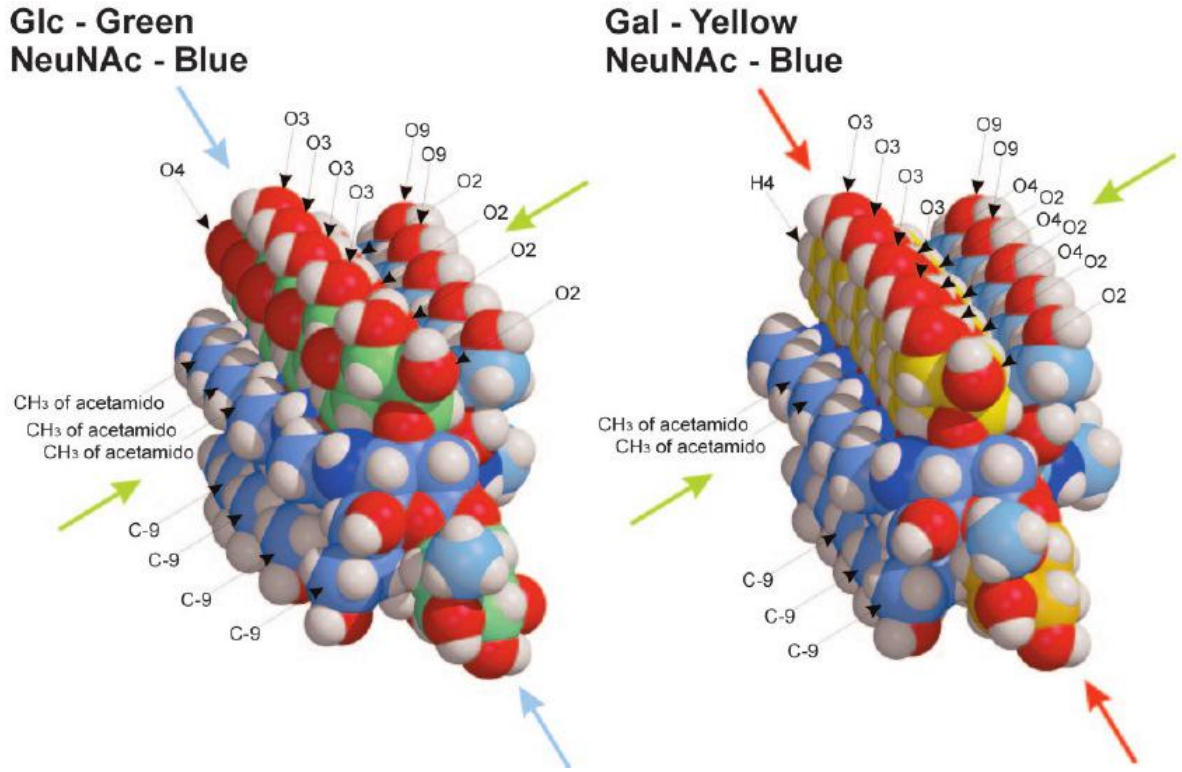


Figure 3.2: Space filling model of 10 RUs of MenY (left) and MenW (right) showing hydrophobic and hydrophilic regions. Hydrophobic surface (green arrow), hydrophilic surface (blue and red arrows). Reproduced with permission from Samuel L. Moore, Catherine Uitz, Chang-Chun Ling et al. (2007), © Clinical and Vaccine Immunology, American Society for Microbiology.[57]

There are several limitations with the Moore *et al* study. The simulations were performed with molecular mechanics, which does not provide a time-based molecular trajectory showing the evolution of molecular conformation with time. Additionally, the force fields employed were not calculated specifically for the carbohydrate molecules used in the study, limiting the accuracy of the results.[124] The simulations were also performed in vacuum and, as a result, they do not take into account solvent interactions and counter ion effects.

Metadynamics and MD simulations can be used to improve upon as well as further investigate the results of the Moore *et al* study. The RUs can be modelled in solution over a time series to provide a trajectory

of molecules. This allows for analysis of conformational dynamics whilst making it easier to identify additional conformational types. The simulations can also be performed in solution using counter ions and the latest class of carbohydrate specific force fields. Studies using MD methods would thus be useful in providing additional insight into the CPS conformation RU of the two meningococcal groups.

In this study we aim to use computational studies to predict the conformations of the MenY and MenW CPS RU. Our systematic procedure as described in Section 3.2 builds from modelling of disaccharide units to simulations of 3RU oligosaccharides in solution. The conformational dynamics and glycosidic linkage angles of the structures are analysed to provide insight into molecular conformation. We model chains of 3RU length as work by Moore *et al.* showed that NMR shifts for 3RUs are comparable to that of longer chains.[57] In addition, they demonstrated complete inhibition of antibody binding for 3RUs in both MenY and MenW and thus structural analysis of the 3RU may provide insight into the observed activity.[57] Furthermore, similar molecular modelling studies on 3RU and 6RU of pneumococcal serotypes 19A and 19F found no significant differences in the conformations and hydrodynamics between the shorter and longer chains.[125]

Figure 3.3 provides a dOA MenY/MenW trisaccharide, which includes the RU structure as used in this study and illustrates the glycosidic linkage dihedral angles. The atom groups defining these angles are given in Section 3.2.

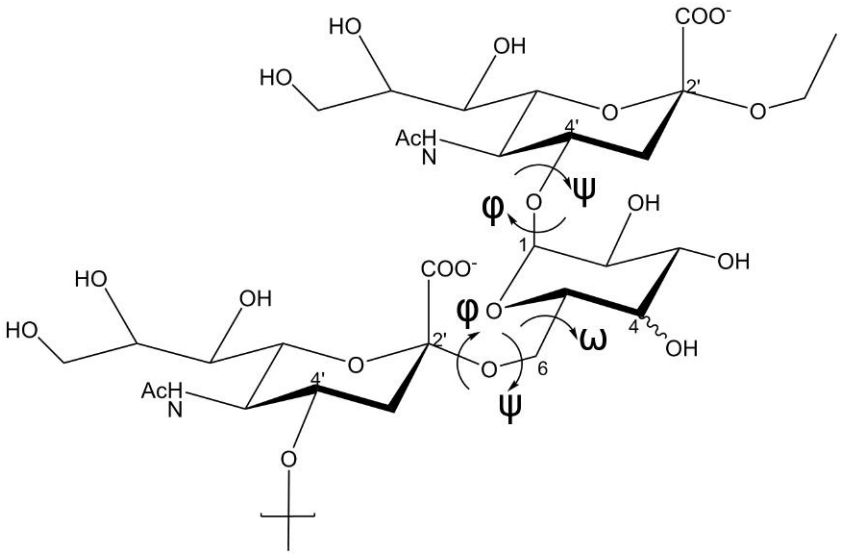


Figure 3.3: dOA MenY and MenW trisaccharide with the repeating unit and illustrating the glycosidic dihedral angles.

Conformation of the hydroxymethyl group

MenY and MenW CPS RU are comprised of 6-linked glucose and galactose pyranose monomers respectively. The rotameric distribution of the hydroxymethyl group has been primarily studied by X-ray crystallography and NMR experiments.[126–128]These studies demonstrated that the conformation of the hydroxymethyl group in glucose and galactose is an equilibrium between three staggered rotameric conformations: gauche-trans (*gt*, ω = +65°), *gauche-gauche* (*gg*, ω = -65°) and trans-gauche (*tg*, $\omega = \pm 180°$) (Figure 3.4).[129]

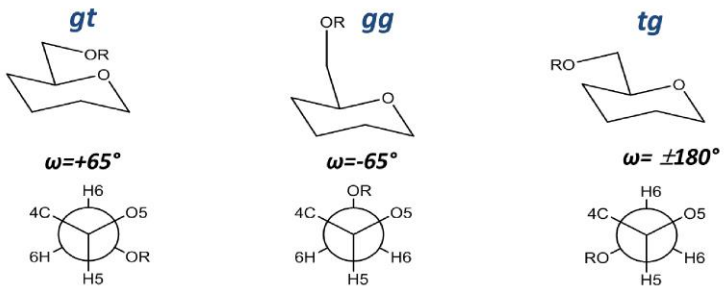


Figure 3.4: Hydroxymethyl conformations of glucose and galactose. In these monomers the *gt* rotamer has no unfavourable interactions and is favoured by the gauche-effect. The *gg* rotamer has unfavourable interactions in galacto-compounds and is favoured by the gauche-effect. The tg rotamer has unfavourable interactions in gluco-compounds and is not favoured by the gauche-effect.[129]

NMR measurement of glucose and galactose monomers in solution identified differences in the rotameric distribution of the hydroxymethyl group. Studies by Nishida *et al.* give a *gg:gt:tg* ratio of approximately 53:45:2 for α-D-glucose and 18:61:21 for α-D-galactose.[126,127] Further studies several years later by Bock *et al* give similar results; a *gg:gt:tg* ratio of approximately 52:41:7 for α-D-glucose and 12:56:32 for α-D-galactose.[128]

There may be several reasons for these observed differences. Glucose has a small population of *tg* rotamers as a result of 1,3 steric interactions between O-6 and the C-4 hydroxyl. This is known as the Hassel-Ottar effect.[130] In addition, for both residues, the *gt* and *gg* rotamers, which have O-6 gauche to O-5, are favoured over the *tg* rotamer, which does not. This may be due to the gauche effect that arises from unfavourable interactions between the anti-bonding orbitals of the C5 – O5 and C6 – O6 bonds in the *tg* rotamer.[129] The force field for glucose and galactose used in this study has been constructed to reproduce the results of these effects.[107–110]

Structure abbreviations and dihedral angle definitions

Table 3.2 provides the glycosidic linkages and corresponding abbreviations as used in this study.

Table 3.2: Structure and abbreviations of meningococcal groups studied.

Structure/Linkage	Abbreviation	Group
α-D-Glc-(1$\rightarrow$4)-α-D-NeuNAc	Glc14N	MenY
α-D-NeuNAc-(2$\rightarrow$6)-α-D-Glc	N26Glc	MenY
α-D-Gal-(1$\rightarrow$4)$-$$\alpha$-D-NeuNAc	Gal14N	MenW
α-D-NeuNAc-(2$\rightarrow$6)-α-D-Gal	N26Gal	MenW

The atoms that form the dihedral angles for the 1$\rightarrow$4 glycosidic linkage in Figure 3.3 are here defined as:

$$\phi = H_1 - C_1 - O_1 - C'_4$$
$$\psi = C_1 - O_1 - C'_4 - H'_4$$

The dihedral angles of carbohydrates bonded through a C-6 linkage as in Figure 3.3 are here defined as:

$$\phi = O'_1 - C'_2 - O_6 - C_6$$
$$\psi = C'_2 - O'_6 - C_6 - H_6$$
$$\omega = O_6 - C_6 - C_5 - O_5$$

3.2 Methods

We performed computational modelling of meningococcal group Y and W CPS RUs in two phases. The first identifies preferred dihedral angle orientations of the glycosidic linkages using metadynamics simulations.[16,125] In the second phase we use these preferred angles to build larger 3RU structures, which were simulated with molecular dynamics in solution.

The initial configurations of the disaccharides and 3RU oligosaccharides were generated with CarbBuilder,[131,132] which uses the psfgen tool[133] to create a 'protein data bank' (PDB) and 'protein structure file' (PSF)[134] for use with Nano Scale Molecular Dynamics (NAMD).[135] The PDB and CHARMM PSF structure files for the starting structures of the molecules simulated are provided in Appendix A.

Phase 1: Potential of mean force calculations

The metadynamics routine incorporated into NAMD v2.9 was used to calculate the ϕ, ψ and ω potential of mean force free energy surfaces for the disaccharide linkages. Simulations were performed on four disaccharides in solution namely: Gal14N, Glc14N, N26Glc, and N26Gal (Table 3.2).

In each simulation the disaccharide was placed in the centre of a cubic box with sides of 6 nm. The box was filled with approximately 32718 water molecules and a single Na^+ counter ion. The Glc14N and Gal14N residues were simulated for 600 ns and the N26Glc and N26Gal disaccharides for 1300 ns. The PMF simulations were stopped when no new minima appeared within 200 ns of formation of the last minimum; and the existing minima did not shift position by more than 5° for at least 100 ns. Structures were collected at intervals of 250 ps for analysis.

A hillWeight of 0.5 was used for Glc14N and Gal14N, and 0.1 for N26Glc and N26Gal. The collective variables were defined as ϕ and ψ for both Glc14N and Gal14N, and ϕ and ω for both N26Glc and N26Gal. The ψ dihedral for N26Glc and N26Gal was not modelled as a collective variable in the metadynamics simulations as the $C_6 - O_6$ bond has been shown to have a single preferred low energy orientation.[136] The NAMD configuration and collective variable files for the metadynamics simulations are provided in Appendix A.

Phase 2: Molecular dynamics of repeating units

MD simulations in solution were performed on a 3RU of MenY and MenW. Using CarbBuilder[131,132], the glycosidic linkage dihedrals of the starting structures were set to initial values within the global energy minima identified by the PMF calculations. The starting structures as obtained from the respective metadynamics simulations were placed in the centre of a cubic box with sides of 12 nm, solvated with

approximately 11 000 water molecules and neutralised with 3 Na^+ ions using the same method as that of the metadynamics simulations. All the MD simulations were run for 200 ns. The system was equilibrated 0.03 ns with a cycled temperature increase from 0 K to 300 K in 10 K increments with each cycle commencing with a 10 000 step energy minimisation followed by a 0.001 ns MD simulation at the specified temperature until 300 K. Structures were collected at intervals of 250 ps for analysis.

General simulation parameters

All metadynamics and MD simulations were performed using NAMD v2.9 for Linux-x86_64 or Linux-x86_64-multicore-CUDA.[135] The latter package employs NAMD CUDA extensions for the calculation of long-range electrostatics and non-bonded interactions on GPUs. The GPU-accelerated simulations were run on a 12 core, 4 TeslaM2090 GPU server with 64GB RAM per core and 6GB DDR5 per GPU. Visualisation and analysis was undertaken with Visual Molecular Dynamics (VMD) v1.9.2[137] or Python scripts. Solvation was performed using the VMD Solvation Box and Ions plugin.[138,139]. Where necessary, molecular conformations were depicted with the PaperChain and Twister visualisation algorithms for carbohydrates[140] to highlight the glycan rings.

The metadynamics and MD simulations were performed in a cubic water box with each dimension approximately 2.5 times the longest length of the carbohydrate molecule. The TIP3P[115] water model was employed and Na^+ counter ions were used to neutralise the solution. All metadynamics and MD simulations were preceded by a standard NAMD 10 000 step energy minimisation protocol pre- and post solvation and ionisation. A modified CHARMM36 additive force field was used.[107–110] The NAMD velocity-verlet integrator[114,135] was employed with a 1 fs time step. Particle mesh Ewald (PME) summation was used for long-range electrostatic interactions with $\kappa = 0.20$ Å^{-1} and PME grid dimensions of 6 nm for the disaccharide simulations and 12 nm for the 3RU simulations. Non-bonded interactions were truncated with a switching function applied between 1.2 nm and 1.5 nm to groups with integer charge. The 1-4 interactions were not scaled as per CHARMM force-field recommendations.

All simulations employed periodic boundary conditions and were performed at 300 K and 1 atm under NPT. Temperature and pressure were coupled using the Nose-Hoover thermostat[141,142] and Langevin piston barostat[143] methods. A NPT ensemble was utilised as opposed to a NVT ensemble as constant pressure is recommended for periodic simulations in NAMD.[144]. The first 10 s of the simulations were regarded as an equilibration phase and were not included in the analysis unless indicated otherwise.

Cluster analysis was done using VMD's internal measure cluster command to calculate clusters based on the quality threshold algorithm.[118]. Structures from the MD trajectory were aligned to the first frame and clustering was performed on the ring and glycosidic atoms of the molecules with a RMSD cut-off size of 0.2 nm. Only clusters with greater than 5% simulation time were analysed. SASA analysis was performed using VMD's internal measure SASA command, which is based on the Shrake-Rupley algorithm.[145,146]. An orb radius of 0.14 nm was used for the SASA analysis.

3.3 Results and discussion

We investigated the MenY and MenW CPS conformations by first studying the disaccharide RU structures obtained from metadynamics simulations and identified dihedral orientations of the low energy structures. We then evaluated the MD simulation results by analysing conformation and dynamics of the MenY and MenW 3RU oligosaccharides chains.

3.3.1 Potential of mean force of calculations of Glc14N and Gal14N

Figure 3.5 illustrates the ϕ and ψ dihedral angle PMF surfaces of Glc14N and Gal14N calculated from the metadynamics simulations as well as representative structural conformations at the minimum regions.

As expected the surfaces are very similar as the sole difference between the dimers is the orientation of the C-4 hydroxyl (equatorial for Glc14N and axial for Gal14N), which is located at the opposite end of the pyranose ring from the 1→4 glycosidic linkage. The PMF surfaces of both disaccharides have three minima regions at ϕ, ψ = -25°, +40° (ΔG = 3 kCal mol^{-1}); ϕ, ψ = -34°, $\pm$180° (ΔG = 5 kCal mol^{-1}); and ϕ, ψ = -33.75°, -41.25° (ΔG = 0 kCal mol^{-1}, the global minimum). Higher energy regions above ΔG = 5 kCal mol^{-1} mostly lie outside of the -90° and 30° range for ϕ and -100° and 90° range for ψ.

The lowest energy conformer (**X** in Figure 3.5) of both Glc14N and Gal14N orientates to allow for a maximum distance between the monosaccharide residues with the ring faces positioned in a non-parallel manner. In addition, the amide substituent of the sialic acid residue for both disaccharides lies above the face of the glucose or galactose ring. Consequently, H-5 lies within 0.35 nm to 0.4 nm of the amide functional allowing for long range hydrogen bonding between the individual monomers.

3.3.2 Potential of mean force calculations for N26Glc and N26Gal

The ϕ and ω dihedral angle PMF surfaces of N26Glc and N26Gal are also very similar, although, there are a several key differences between them. (Figure 3.6). The surfaces each contain 9 minimum regions with the global minimum for N26Glc at **c** (ϕ, ω = 61°, 64°) and the global minimum for N26Gal at **b** (ϕ, ω = 64°, 59°). Additional minima lie at positions **a** and **d**-**i**. The dihedrals and free energies for key minimum points are summarised in Table 3.3. The free energies at minima ϕ, ω = -36°, -64° (**f**) and ϕ, ω = $\pm$180°, -64° (**i**) are the same for both N26Glc and N26Gal.

Table 3.3: Values of the ϕ, ω dihedrals for minima labelled on the PMF surfaces of N26Glc and N26Gal. The free energies (in kCal mol^{-1}) relative to the global minimum of each disaccharide are shown in brackets.

Disaccharide	**Conformational region**									
	c		**f**		**i**		**b**		**a**	
	ϕ, ω	(ΔG)	ϕ, ω	(ΔG)	ϕ, ω	(ΔG)	ϕ, ω	(ΔG)	ϕ, ω	(ΔG)
N26Glc	61°, -64°	(**0**)	-36, -64°	(4)	$\pm$180°, -64°	(5)	64°, 59°	(1.4)	56°, 174°	(3.5)
N26Gal	61°, -64°	(0.4)	-36, -64°	(4)	$\pm$180°, -64°	(5)	64°, 59°	(**0**)	61°, 177°	(1.3)

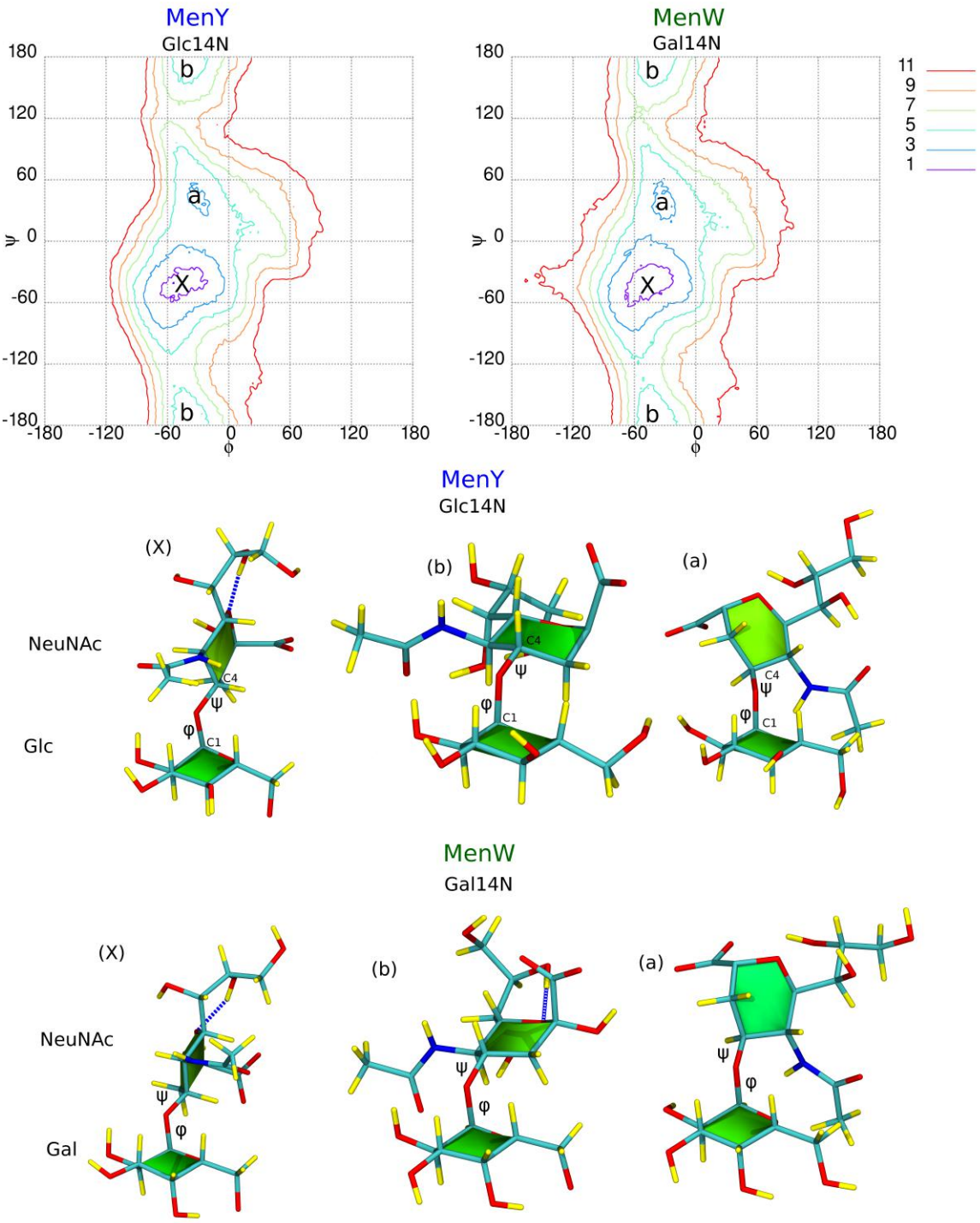


Figure 3.5: Solution PMF surfaces of αDGlc(1→4)αDNeuNAc (top left) and αDGal(1→4)αDNeuNAc (top right). Labels **X**, **a**, and **b** denote the minima regions for both structures. Contours lie at 2 kCal mol^{-1} increments with the energies relative to the the global minimum at **X**. Below are representative structures of Glc14n and Gal14N at positions **X** and **b** on their respective PMF surfaces. Dashed blue lines show hydrogen bonds.

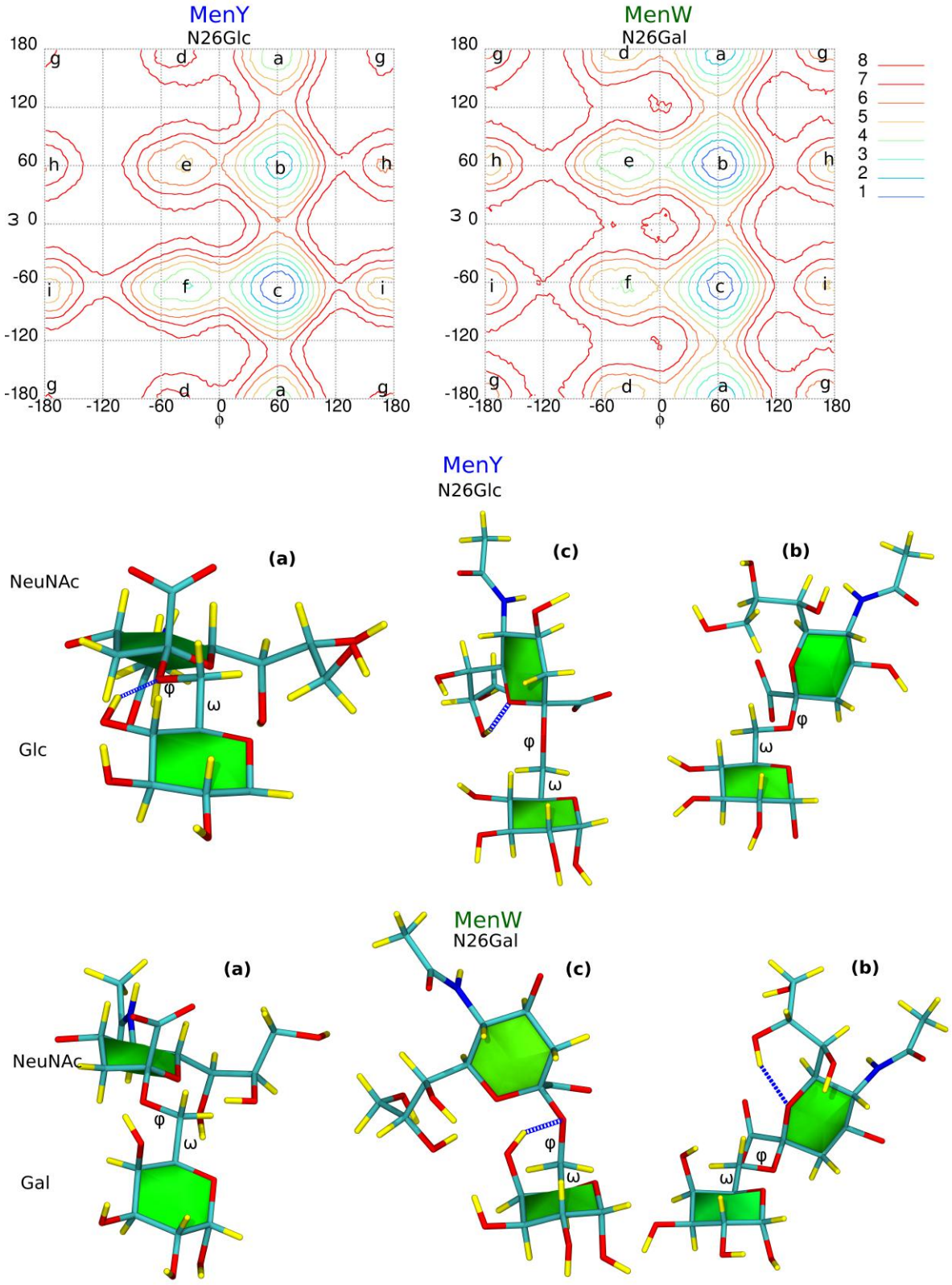


Figure 3.6: MenY and MenW PMF plots in solution for αDNeuNAc(2→6)αDGlc (left) and αDNeuNAc(2→6)αDGal (right). Contours are drawn at 1 kCal mol^{-1} intervals. The global minimum is at **c**; notable local minima lie at **a**, **b**, **i** and **f**. Representative structures of N16Glc and N16Gal (below) showing the three significant ω dihedral rotameric conformations. The *tg* rotamer (left), the lowest energy *gg* rotamer (centre) and the *gt* rotamer (right). Dashed blue lines show hydrogen bonds.

The PMF surface profile of N26Glc and N26Gal is closely related to the rotameric conformation of the hydroxymethyl group in glucose and galactose.[119] In glucose residues the *gt* (ω = 65°) and *gg* (ω = -65°) rotamers typically have no unfavourable 1,3-interactions and are favoured by the gauche-effect. In contrast, the *tg* ($\omega = \pm 180°$) rotamer has unfavourable 1,3-interactions in gluco-compounds and is not favoured by the gauche-effect. The PMF surface of N26Glc reflects these differences. The *gg* rotamer (**c**) is lowest energy (ΔG = 0 kCal mol^{-1}) while the *tg* rotamer (**a**) is highest in energy (ΔG = 3.5 kCal mol^{-1}). The *gt* rotamer (**b**) is 1.4 kCal mol^{-1} higher than the global minimum and this difference in energy may be attributed to steric effects.

In galactose residues *gt* and *tg* rotamers are typically preferred whilst the *gg* form are less so due to unfavourable 1,3-interactions. In the PMF surface of N26Gal the *gt* rotamer (**b**) is lowest in energy, while the *gg* (**c**) and *tg* ((a) rotamer are 0.4 kCal mol^{-1} and 1.3 kCal mol^{-1} higher in energy respectively. This result shows that even though the minima of N26Glc and N26Gal lie on the same regions on their PMF surfaces there are significant differences in the free energies at these points. These differences in preferred glycosidic linkage confirmation may have a significant effect on polymer conformation.

3.3.3 Polymer conformations

We provide here an analysis of the conformation and dynamics of the 3RU oligosaccharides of MenY and MenW. We studied orientation of the glycosidic linkages, molecular folding, hydrogen bonding, solvent accessible surface area, and cluster conformations. Additional analysis of root mean square deviation, radius of gyration, dihedral measurements, and ring puckering from the MD simulations are given in Appendix A. The PDB and PSF files of the starting structures are also listed in Appendix A. The chemical formulae of the MenY and MenW 3RU simulated as well as labels identifying the glycosidic linkages are provided below:

i **ii** **iii** **iv** **v**

Y: α-D-Glc(1→4)-α-D-NeuNAc-(2→6)-α-D-Glc(1→4)-α-D-NeuNAc-(2→6)-α-D-Gal(1→4)-α-D-NeuNAc

W: α-D-Gal(1→4)-α-D-NeuNAc-(2→6)-α-D-Gal(1→4)-α-D-NeuNAc-(2→6)-α-D-Gal(1→4)-α-D-NeuNAc

Glycosidic linkage conformations

Figure 3.7 shows the Glc14N/Gal14N and N26Glc/N26Gal dihedral angles of the MenY and MenW 3RU strands over the course of the 200 ns simulation. The plots for all three Glc14N and Gal14N linkages (**i**, **iii** and **v**) show that the ϕ and ψ dihedrals remain close to the global minimum energy well. Most lie within 1 kCal mol^{-1} of the global minimum, **X**, near ϕ, ψ = -33.75°, -41.25°, while almost all fall within at least 2 kCal mol^{-1} of this point. A very small number of the measured dihedrals are also located in local minima at **a** and **b**.

In contrast, the N26Glc and N26Gal plots in Figure 3.7 show clear differences in dynamic behaviour. The N26Glc linkages remain almost exclusively within the global minimum energy well (**c**), with minima **b** and **f** only very slightly populated. The remaining energy minima are not significantly populated. This suggests that the *gg* rotamer is the sole preferred conformation of the N26Glc linkage.

The N26Gal plot shows a significantly different profile. The linkage also remains primarily in minimum

energy well **c** but in addition shows frequent rotations about the ω dihedral populating minima **a** and **b**. The dihedrals also briefly populated minima **f** and **i**, while the remaining minima are not visited. The distribution of the N26Gal linkage dihedrals indicate that the *gg* rotamer is preferred, although, both the *gt* and *tg* forms are accessed for significant periods during the simulation.

These results closely correspond to the free energies measured for N26Glc and N26Gal. The barriers of the three lowest energy wells (**a**, **b** and **c**) of N26Gal are between 1 kCal mol^{-1} and 2 kCal mol^{-1} lower in energy than the corresponding energy barriers of N26Glc. Furthermore, the local minima of N26Gal also lie lower in energy than the local minima for N26Glc. These results are reflected in the MD simulations by the population of the different rotameric states of the 3RUs. The N26Gal linkages of the MenW 3RU are able to move over the lower energy barriers between and populate the *gg*, *gt* and *tg* minimum regions. However, the N26Glc linkages of the MenY 3RU are not able overcome these energy barriers and remain almost entirely in the *gg* rotameric conformation.

To further study dynamic behaviour of the glycosidic linkages we analysed dihedral angle probability distributions. The ϕ and ω probability density functions for the N26Glc and N26Gal linkages are provided in Figure 3.8 and 3.9 respectively.

The distribution of the ϕ dihedral for the N26Glc and N26Gal linkages are similar. For the N26Glc linkages, ϕ lies between -50° and -100° for over 95% of the structures in the MD trajectory. The ϕ dihedral of the N26Gal linkages have a similar distribution, with approximately 93% of the dihedral angle in the 50° to 100° range. A region near -40° is also slightly more densely populated in the linkage **ii** of N26Gal compared to the same region in the N26Glc linkage. The density function of the ω dihedrals for the N26Glc and N26Gal linkages clearly show the differences in the angle distributions of MenY and MenW 3RU. The N26Glc density function contains a single peak between -40° and -100° with over 95% of the ω dihedral falling within this range. Similarly, the single largest peak for the N26Gal linkages also falls within this range although the size of the region is significantly smaller at about 60%. The two

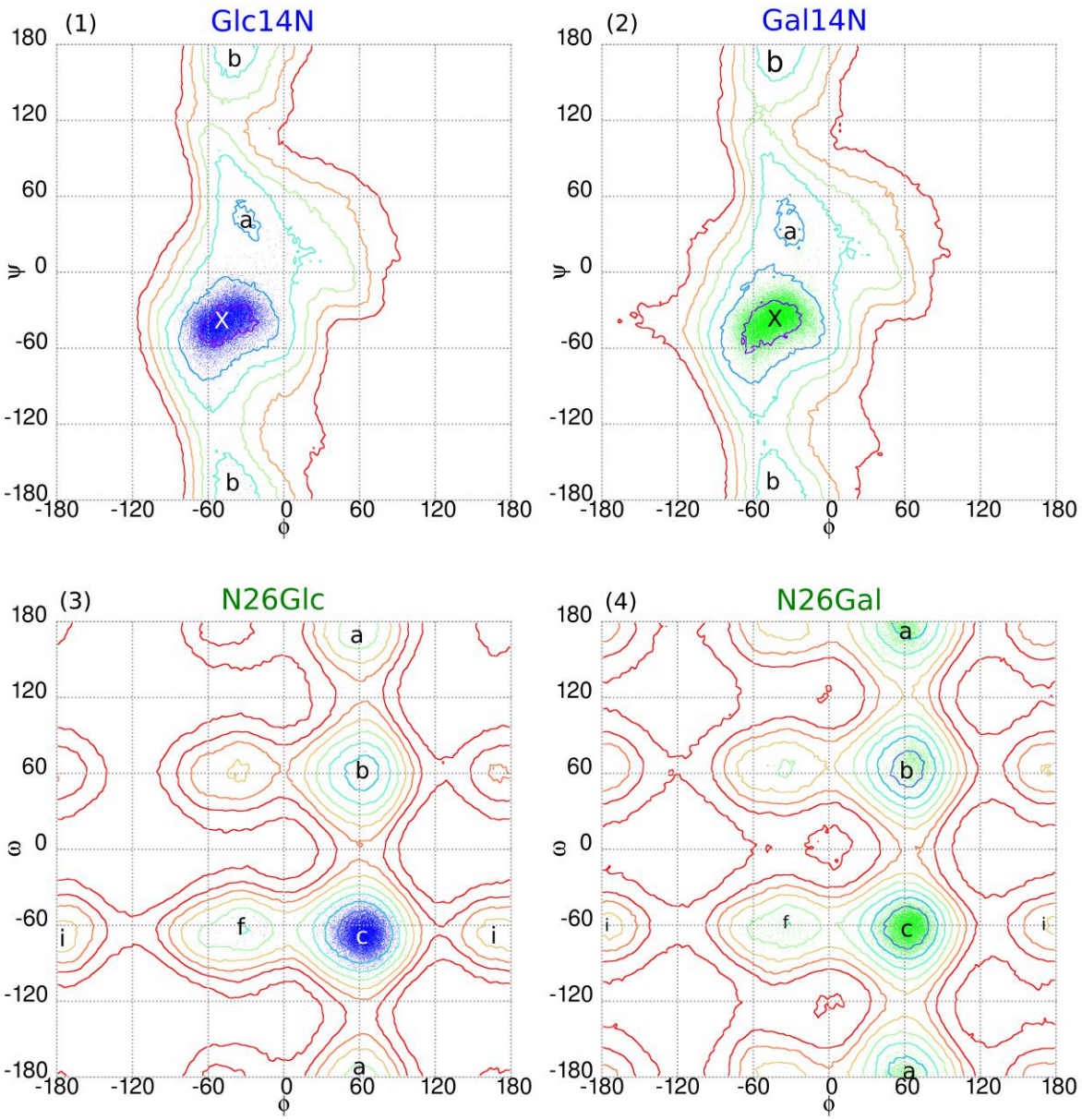


Figure 3.7: Overlay of the MD dihedral angles of MenY (left) and MenW (right) 3RU structure on metadynamics free energy surfaces. (1) and (2) provide the dihedral angles for linkages **i**, **iii** and **v** for Glc14N and Gal14N respectively. (3) and (4) provide the dihedral angles for linkages **ii** and **iv**. Contours lie at 1 kCal mol^{-1} increments.

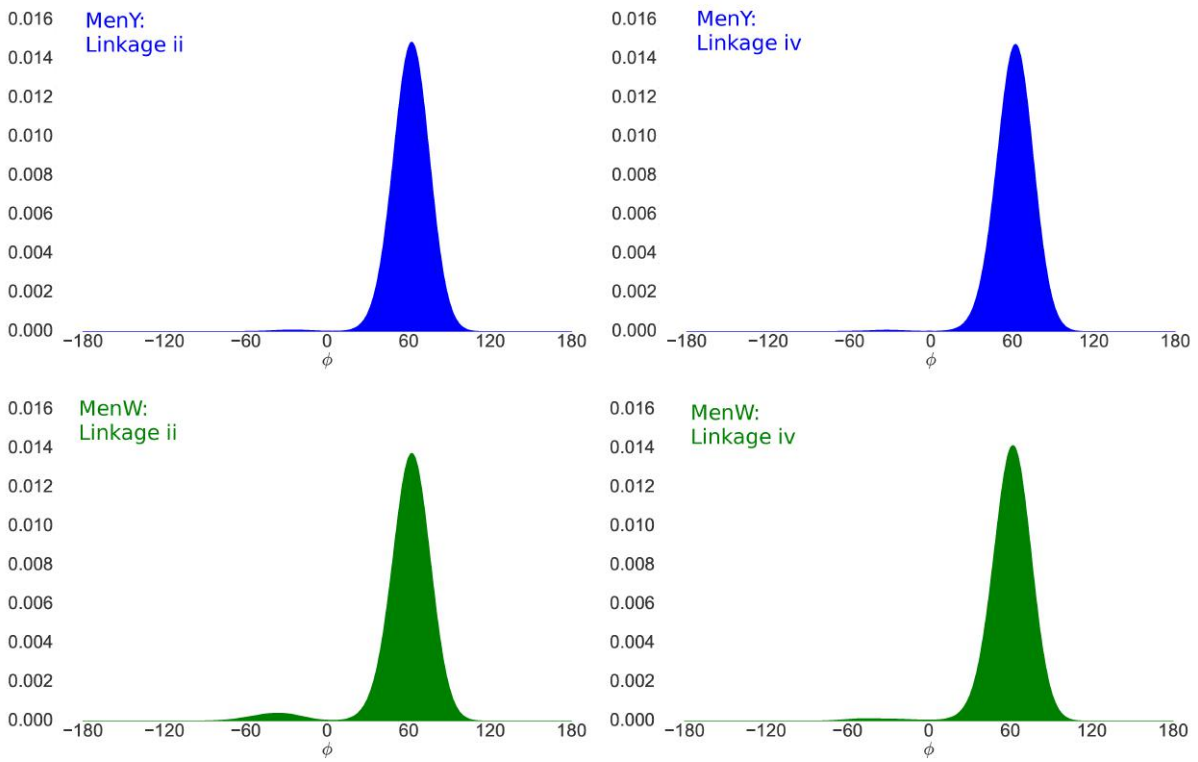


Figure 3.8: Probability density functions of the N26Glc (above) and N26Gal (below) linkage ϕ dihedrals in the MenY and MenW 3RU.

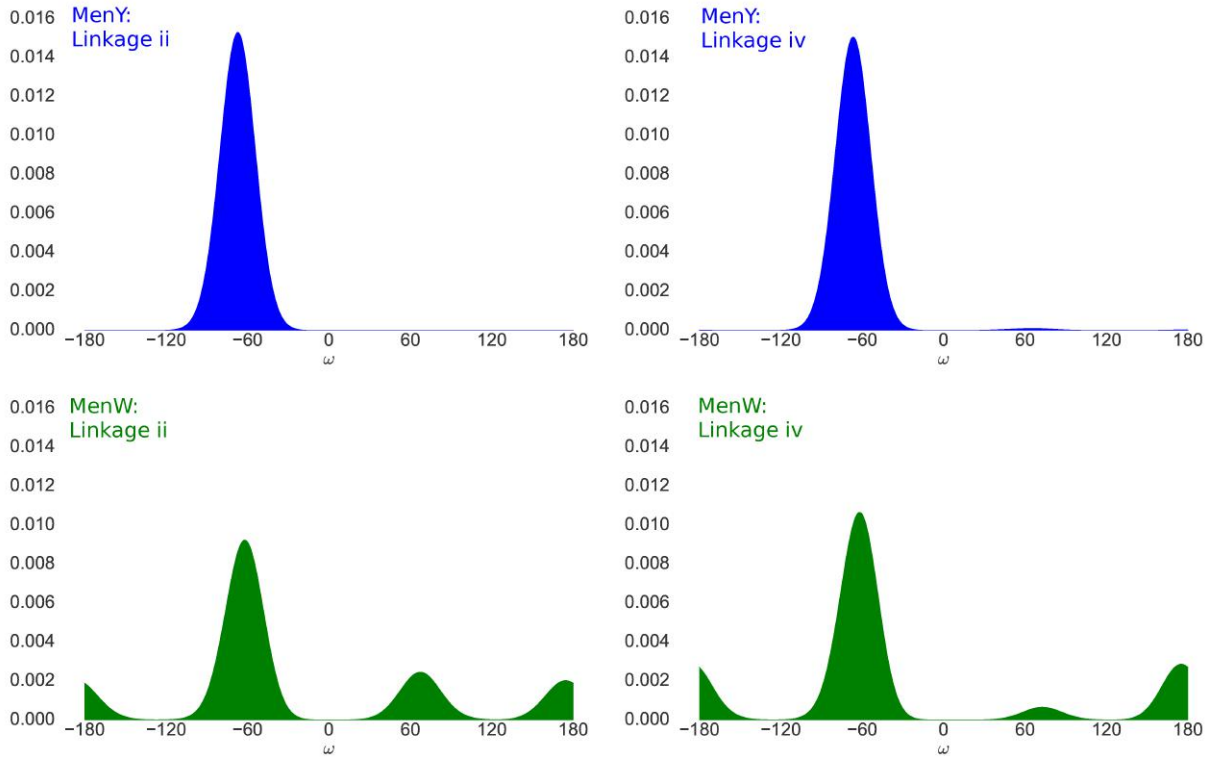


Figure 3.9: Probability density functions of the N26Glc (above) and N26Gal (below) linkage ω dihedrals in the MenY and MenW 3RU.

smaller peaks constitute the remaining populated dihedral angles with about 20% between -150° and 150° and 15% between 40° and 100°. The preferred dihedrals for the 1→4 and 2→6 linkages of MenY and MenW 3RU are summarised in Table 3.4.

The density functions of the N26Glc linkage ω dihedrals confirms the *gg* rotamer in the MenY 3RU is preferred. This results corresponds to the standard glucose model for rotameric conformation. In the galactose model the *tg* rotamer is also the preferred rotamer, however, the simulated structure for N26Gal has the *gg* rotamer as the most prevalent followed by the *tg* and *gt* forms. This suggests that the unfavourable 1,3 steric interactions between OH-4 and C-6 of the galactose residue is not the dominant factor in determining dihedral orientation and that one or more interactions play a greater role.

Conformational dynamics

To study conformational dynamics we calculated the change in the end-to-end distance‡ for the MenY and MenW 3RU. We also measured the ω dihedral of the N26Glc and N26Gal linkages and correlate dihedral orientation to the end-to-end distances. The plots in Figure 3.10 illustrate the change in end-to-end distance and ω dihedral of the MenY and MenW 3RU from their respective 200 ns MD simulations.

The end-to-end distance of the MenY 3RU (Figure 3.10-1) ranges from 1.2 nm to 2.4 nm, although, the 10 ns average distance remains relatively constant at 1.9 nm with a standard deviation of 0.2 nm. There are marked changes in the end-to-end distance of the MenW 3RU (Figure 3.10-2) during the MD simulation. The distance changes little in the first 80 ns of the simulation with an average between 1.9 nm and 2 nm. For 30 ns after this the average end-to-end distance increases to over 2.2 nm and then falls to 1.8 nm for the next 40 ns. In the remaining portion of the simulation, the MenW 3RU is distinctly flexible with the 10 ns average end-to-end distance moving consistently between 1.8 nm and 2.2 nm.

The time series of the N26Glc linkage ω dihedrals illustrates their single preferred orientation in the MenY 3RU. In contrast, the behaviour of the ω dihedrals in the MenW 3RU is significantly different. The ω dihedral for both the MenY and MenW 3RU are close to -60° at the start of the simulation. The N26Glc linkages oscillate around this value throughout the simulation with the exception of a short

Table 3.4: Preferred dihedral angles and population sizes based on the probability distributions of MenY and MenW 3RU. The 1→4 linkages have no ω dihedral.

Group	Dihed.	α-D-G*(1→4)-α-D-NeuNAc-(2→6)-α-D-G*(1→4)-α-D-NeuNAc-(2→6)-α-D-G*(1→4)-α-D-NeuNAc				
		i	ii	iii	iv	v
Y	ϕ	-50°, >95%	62°, >95%	-50°, >95%	63°, >95%	-49°, >95%
	ψ	±180°, >95%	-172°, >95%	±180°, >95%	-173°, >95%	±180°, >95%
	ω	-	-67°, *gg*, >95%	-	-65°, *gg*, >95%	-
W	ϕ	+48°, >95%	62°, >95%	+48°, >95%	62°, >95%	50°, >95%
	ψ	±180°, >95%	-172°, >95%	±180°, >95%	-171°, >95%	±180°, >95%
	ω	-	-62°, *gg*, 60%	-	-63°, *gg*, 60%	-

‡ The end-to-end distance is the distance between the ring oxygen atoms on the terminal residues of the MenY and MenW 3RU

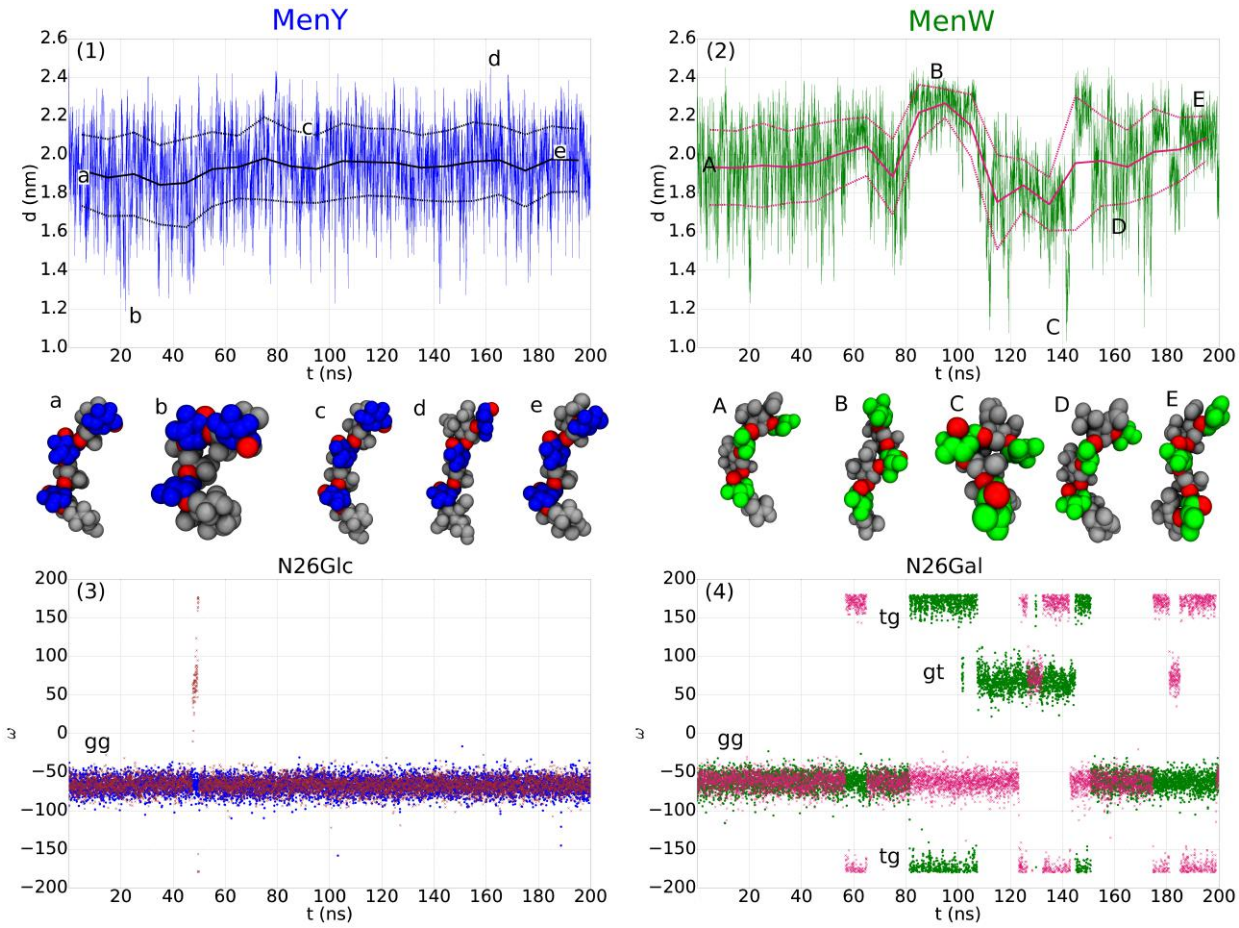


Figure 3.10: (Top) Time series of end-to-end distances from MD simulations of 3RU of MenY (**1**) and MenW (**2**). Centre lines show 10 ns mean, outer lines show 10 ns standard deviations. Distance measured from ring oxygen atoms of terminal residues; (Bottom) Time series of glycosidic linkage ω dihedrals from MD simulations of 3RU of MenY (**3**) and MenW (**4**). Linkage **ii** represented by (·) and **iv** by (x).

period between 47 ns and 50 ns where linkage **iv** changes to +60° and ±180° after which it rapidly moves back to the -60° range.

The ω dihedrals of the N26Gal linkages in the MenW 3RU remain close to -60° for the first 60 ns in the simulation. For the remaining 140 ns ω dihedral of either one or both linkages moves between -60°, +60° and ±180°. The change in the dihedral angle closely correlate to the MenW 3RU end-to-end distances during this period; with the end-to-end distance rising above the average when one or both the dihedrals are at ±180° or below the average when the dihedrals are near +60°.

The graphs show correlation between the end-to-end distances and the the ω dihedral. In the MenY 3RU, both the distance and ω dihedral remain constant throughout the simulation while in the MenW 3RU both fluctuate significantly. The conformation of structures (**a**-**e**) reflect these differences. The MenY 3RU structures, with the exception of **c** are conformationally similar adopting a helical conformation. In contrast, the MenW 3RU adopts multiple structural types and are significantly more conformationally flexible. These include helical conformations (**A** and **D**), which are similar to that of the MenY 3RU, more folded conformations (**B** and **E**); as well as conformations that are highly compact forming a significantly smaller structural unit (**C**).

Hydrogen bonding analysis of the MD trajectories identified no hydrogen bonds in the MenY 3RU. The MenW 3RU showed two consistent hydrogen bonds between the C-4 hydroxyl of galactose and the glycosidic oxygen of both N26Gal linkages (**ii** and **iv**). The hydrogen bonds were confirmed by measuring the atom-to-atom distance from OH-4 to O-6 as well as the O4 – OH4 – O6 bond angle. We compare this distance and angle for both N26Glc and N26Gal in the MenY and MenW 3RU respectively (Figure 3.11).

The OH-4 to O-6 atom-to-atom distance for both N26Glc linkages remain within 0.3 nm and 0.45 nm

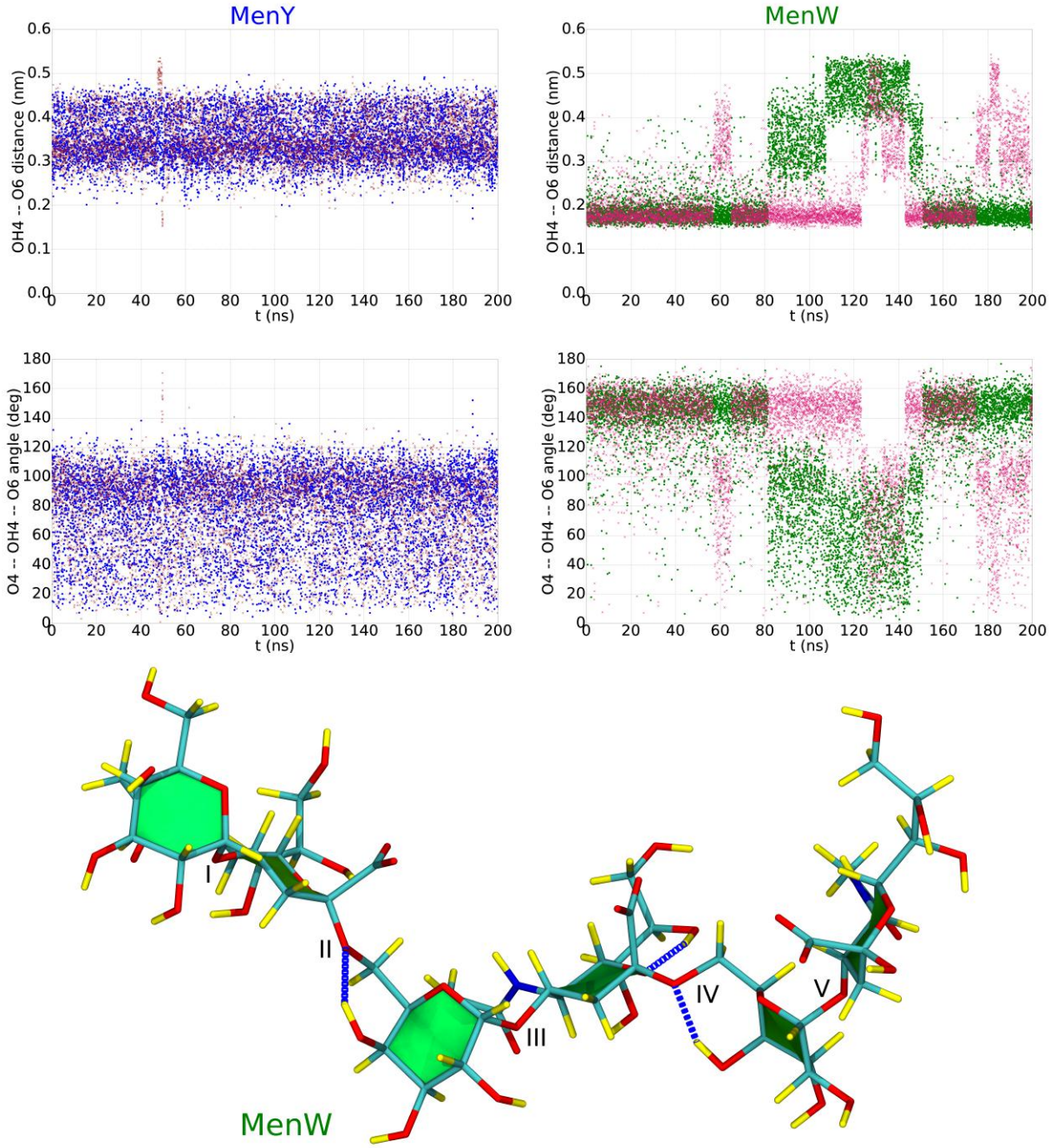


Figure 3.11: (Above) Time series of OH4 – O-6 distance and O4 – OH4 – O6 angle of MenY (left) and MenW (right) 3RU. (Below) Representative MenW 3RU showing the N26Gal linkages (**ii** and **iv**) in the *gg* rotameric form and the hydrogen bonds between OH-4 and O-6.

for most of the simulation while the O4 – H-4 – O6 angle ranges from 20° to 140° with an average of approximately 90°. The same atom-to atom distance measurements for the N26Gal linkages lie between 0.15 nm and 0.2 nm for over 60% of the simulation. The distance increases up to a 0.55 nm for periods before returning down the 1.5 nm to 2 nm range. Overall, the measured distances confirm that it is highly unlikely a hydrogen bond will form between OH-4 and O-6 in the N26Glc linkages.

For the N26Gal linkages in the MenW 3RU the H-4 to O-6 atom-to-atom distance fluctuates significantly. In the over 140 ns of the simulation the average distance averages below 0.2 nm. During these periods the O4 – H-4 – O6 angle lies mostly above 140°. This average distance and angle would suggest the formation of a hydrogen bond.

Hydrogen bond formation strongly correlates with the N26Gal linkage ω dihedral orientation. The hydrogen bond is present when the ω dihedral is in the *gg* rotameric form. As *gg* is not typically favoured in galactose residues due to 1-3 steric interactions it is likely that the this configuration in the N26Gal linkage is stabilised by the presence of the hydrogen bond.

Solvent accessible surface area

We measure solvent-accessible surface area of the MenY and MenW 3RUs and study correlation between molecular conformation and solvent-accessible area. Figure 3.12 depicts the change in the SASA of the MenY and MenW 3RUs over the course of the MD simulations.

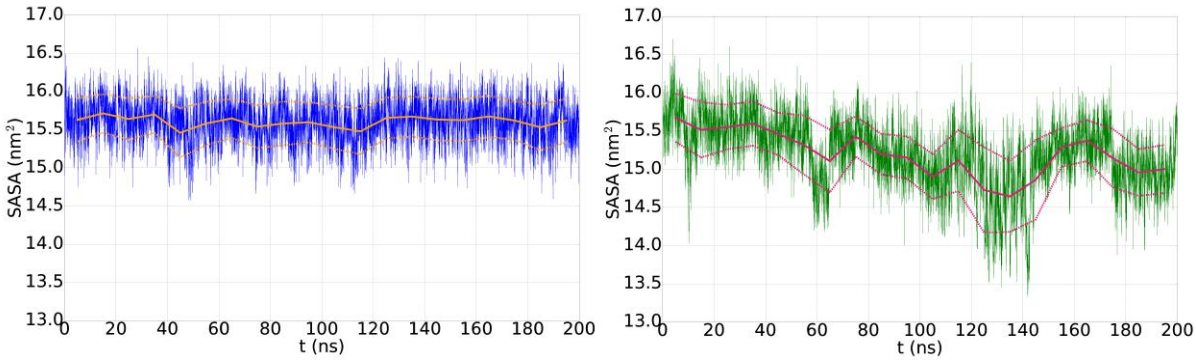


Figure 3.12: Change in solvent accessible surface area of MenY (left) and MenW (right) 3RU over 200 ns of MD in solution. The line and dashed lines are the 10 ns mean average and sample standard deviations respectively

The SASA for the MenY 3RU consistently fluctuates between 14.5 nm^2 and 16.5 nm^2 with an average that remains relatively constant at 15.5 nm^2. The SASA of the MenW 3RU changes over a larger range between 13.3 nm^2 and 16.5 nm^2 while the average also varies notably. The SASA for both 3RU CPSs closely correlates to changes in their N26Glc or N26Gal ω dihedral. The 3RU conformations are similar when the N26Glc and N26Gal linkages are in the *gg* rotameric conformation. When one of the N26Gal linkages changes to either the *gt* or *tg* form folding of the molecule occurs, resulting in a decrease in solvent accessibility. Figure 3.13 provides the molecular surface of representative MenY and MenW 3RU showing a map of the hydrophobic and hydrophilic regions.

The SASA potential maps on the 3RUs show that the dominant structure of MenY as shown on the left of Figure 3.13 has a relatively large number of regions that are solvent inaccessible. The preferred con-

formation of MenW in the centre of Figure 3.13 has a similar SASA profile. However, the conformation at 140 ns on the far right has the 2→6 linkages in the *gg* rotameric conformation and has fewer solvent accessible areas. The total SASA for the structures at 14 ns and 140 ns is 15.85 nm^2 and 14.44 nm^2 respectively. There does not appear to be a large difference in the total solvent-accessible area, however, the second structure shows a deep pocket that is solvent inaccessible.

The dominant conformations of the MenY and MenW CPS RU have similar solvent accessible areas but differ around the hydrophilic C-4 hydroxyl of the glucose and galactose residues. This region in the MenY 3RU is slightly more solvent accessible than the same area in the MenW 3RU. However, the SASA profile of the less dominant MenW 3RU conformations are significantly different. As expected, the SASA decreases with increased conformational folding as a result of formation of solvent occluded pockets.

The MenY and MenW 3RU structures displayed measurable differences in SASA depending on conformation. The dominant conformers for both MenY and MenW have similar SASAs, although, the less preferred conformations of MenW have distinctly different solvent accessible profiles.

Cluster Analysis

We performed a cluster analysis on the MenY and MenW 3RU MD trajectory to identify groups that have similar conformations. The structures were clustered on RMSD with respect to the starting structure and

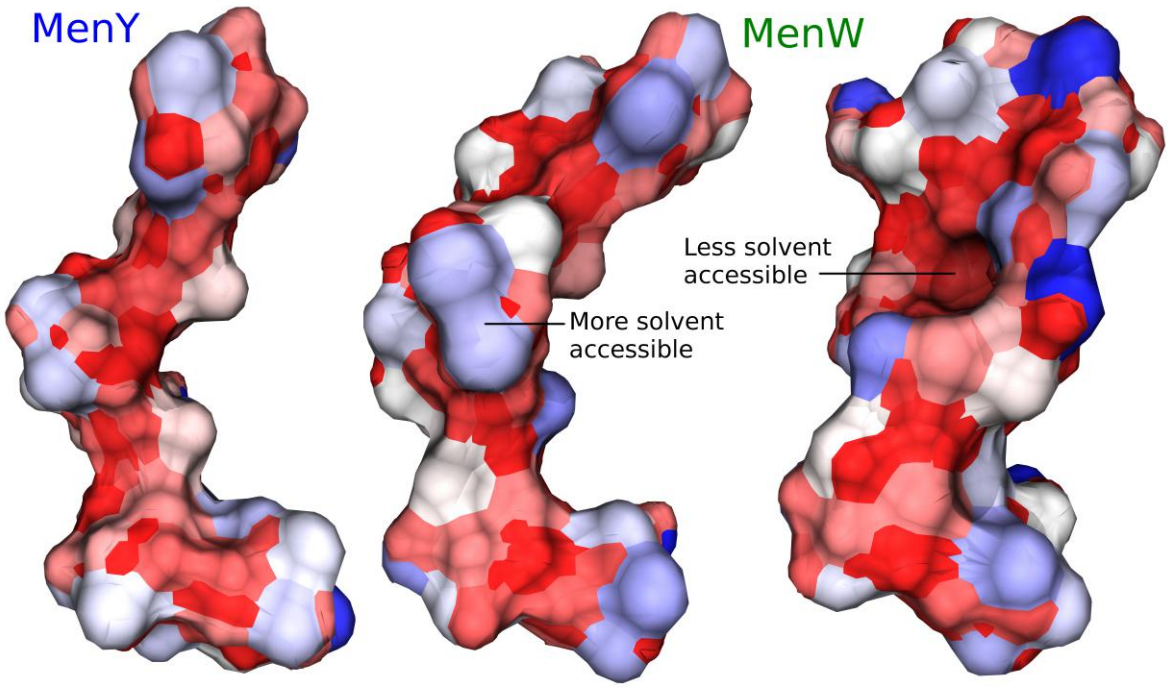


Figure 3.13: SASA potential map on the surface of MenY and MenW 3RU at 14 ns (left and centre), with the N26Glc and N26Gal linkages in the *gg* rotameric form, and for MenW at 140 ns (right), with the N26Gal linkages in the *tg* form. Regions in blue are more solvent accessible while those in red are less solvent accessible.

based on the method in Section 3.2. Figure 3.14 depicts the dominant cluster families of the MenY and MenW 3RUs simulated.

A single cluster was identified from the MD trajectory of the MenY 3RU, with over 95% of structures falling within this group. All the structures in this group have both N26Glc linkages in the *gg* rotameric form.

A total of 3 large clusters and 2 smaller cluster groupings were identified for the MenW 3RU. The conformations of the cluster groups strongly correlate with the rotameric forms of the N26Gal linkages. The largest cluster contains approximately 44% of the structures from the molecular trajectory and both N26Gal linkages in this group are in the *gg* rotameric form.

The second and third largest cluster groups consists of one of the N26Gal linkages in *gg* whilst the other is in the *tg* form. The clusters effectively represent a single conformational group combined with a population size of 27%. The remaining two groups each have one of the N26Gal linkages in the *tg* or *gt* forms and together constitute approximately 15% of the total number of structures.

The conformational structures of the dominant structural families observed are similar to those identified by Moore *et al.*[57] They found that the dominant conformations of MenY and MenW have a related helical structure in the lowest energy conformation, with four residues making one full turn. We also found that the dominant conformation of the MenY and MenW 3RU forms a helical structure with four to five residues forming a single turn. The smaller cluster groups of MenW have a higher degree of folding, which corresponds to the rotameric orientation of the N26Glc linkages, and do not appear to adopt a typical helical shape. Clusters with conformations that have one linkage in the *gt* or *tg* form are partially folded, while the cluster with both N26Gal linkages in the *gt* or *tg* form is more highly compact.

As the relationship between CPS conformation and immunogenic activity has been previously shown,[16,123] these results suggest that there is a correlation between CPS conformation and the level of antigen cross-protection.

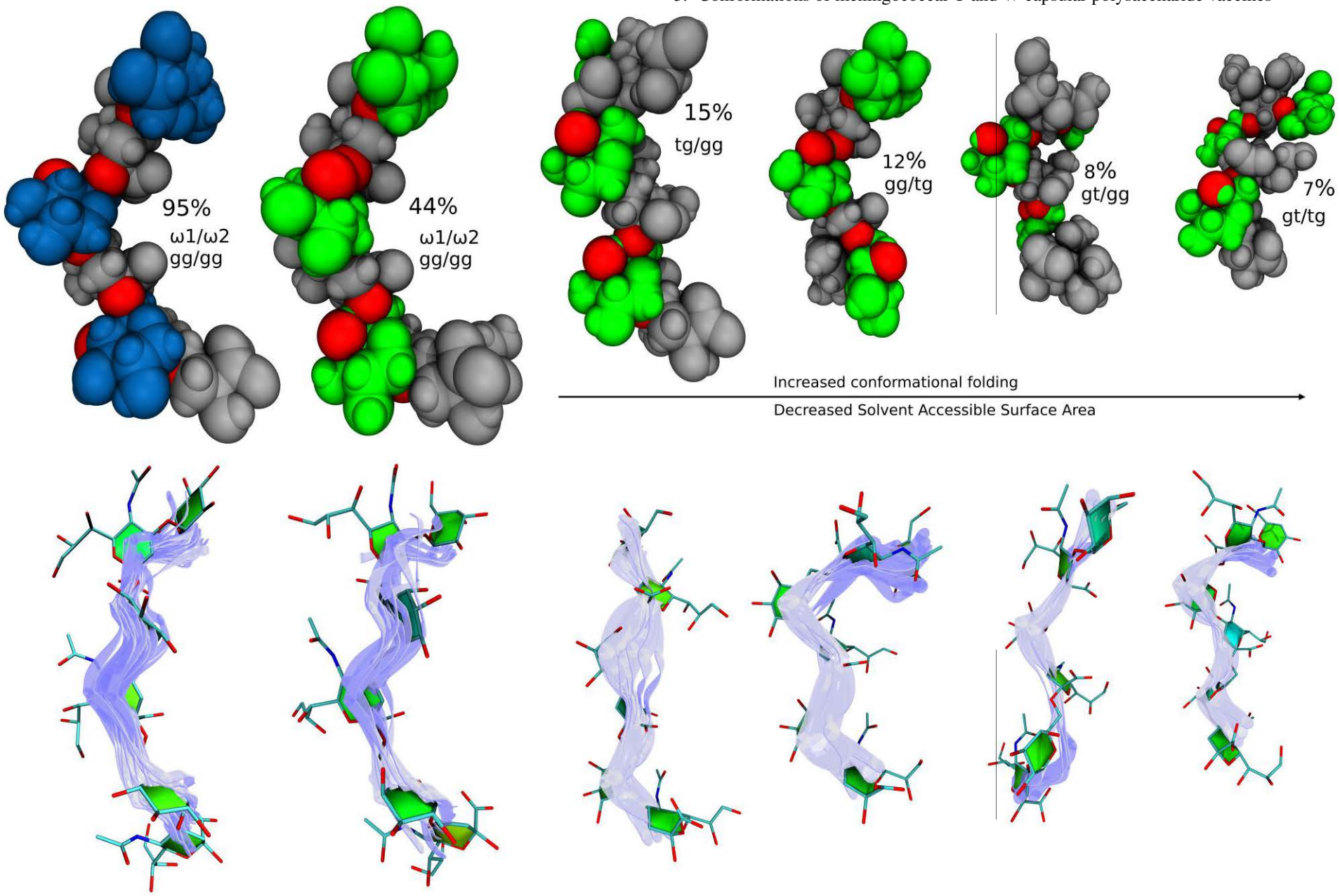


Figure 3.14: Above: Mean representative structures of the MenY (blue) and MenW (green) 3RU structures. The sialic acid residues are grey and oxygen atoms red. The glucose and galactose monomers are coloured. Below: Structural overlays of clusters of selected frames within each cluster grouping. H atoms not shown.
The scale of the smaller clusters has been reduced to emphasize their relative population sizes.

3.3.4 Capsular polysaccharide conformation and immune response

A small scale clinical trial by Griffis *et al.* showed that MenY and MenW CPS vaccines provide differing levels of immune cross-protection (Table 3.1).[89] The study showed that the MenY CPS vaccine was able to elicit high levels of bactericidal antibody against group W bacteria while the MenW CPS vaccine was only able to elicit relatively low levels of bactericidal antibody against group Y bacteria.

The results from our study suggests a correlation between CPS conformation and MenY and MenW vaccine cross-protection. Computational modelling suggests that the MenY CPS adopts a single preferred conformation. Consequently, B-cell receptor recognition of the MenY CPS vaccine antigen will possibly induce the formation of a single antibody type. The antigen binding site of this antibody is specific to the epitope of the dominant MenY CPS conformation.

In contrast, the MenW CPS can adopt multiple conformations, with some conformations that are more dominant than others. Therefore, BCR recognition of MenW CPS vaccine antigen may induce the formation of multiple types of antibodies. Each of these antibodies will be specific to a particular MenW CPS conformation. Additionally, the number of each antibody type formed may be proportional to the different conformational groups of the MenW CPS. Figure 3.15 illustrates a proposed immune response to MenY and MenW CPS vaccines antigens and the associated antibody formation.

These assumptions suggest that, as the dominant conformations of both MenY and MenW CPS are very similar, the antibodies produced from these conformations may also be closely related. This could allow for some antibody cross-protection as antibodies of one meningococcal group should have strong avidity for the dominant CPS conformation of the other group. This is observed in the vaccine trial as both MenY and MenW show antibody cross-protection but at different levels.[89]

Additionally, the results from the vaccine trials suggest differences in the observed level of antibody cross-protection could be attributed to the distribution of antibodies formed by the MenY and MenW vaccine antigens. The MenY vaccine was able to elicit high levels of bactericidal antibody against MenW bacteria. This indicates that the single type of antibody formed by the MenY vaccine antigen may have sufficient avidity with one or more MenW CPS conformations to allow for cell recognition.

In contrast, MenW vaccine was found to have weak levels of antibody cross-protection. The most abundant antibody formed by the MenW vaccine antigen should have strong avidity with the MenY CPS due to the similarity in CPS conformation. However, as the dominant MenW CPS is accessed less than half the time, the number of antibodies formed that are able to cross-protect are significantly less than formed by the MenY vaccine antigen. This suggests that the MenW vaccine has weaker levels of cross-protection as, for a fixed amount of vaccine, it generates fewer antibodies that are able to cross-protect when compared to the MenY vaccine.

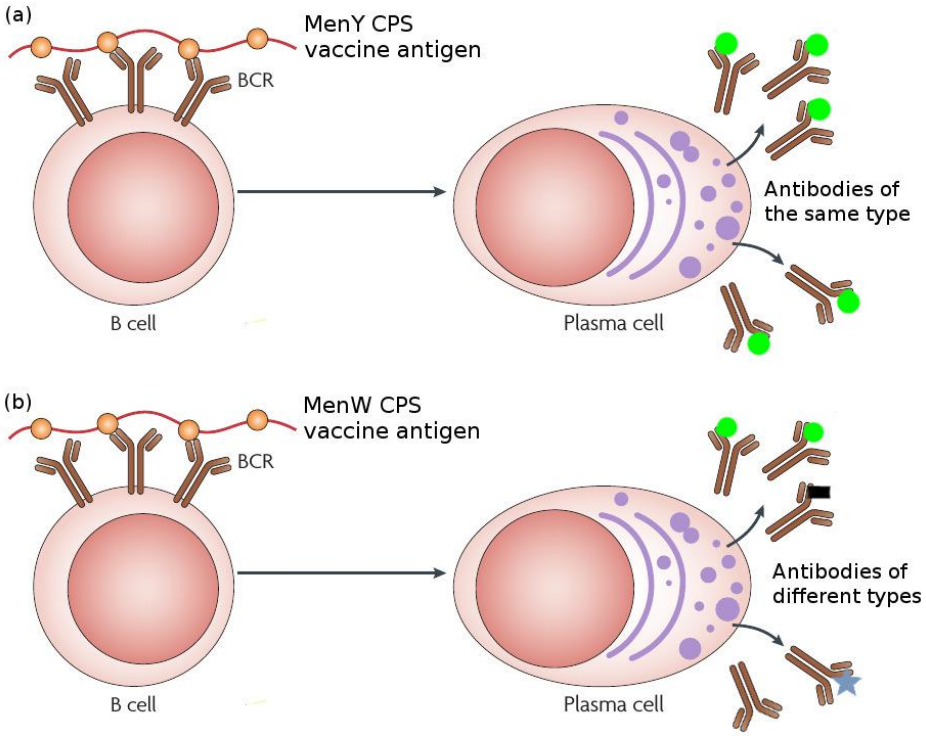


Figure 3.15: Proposed immune response to MenY and MenW vaccine antigens that may lead to different levels of antibody cross-protection. (a) B-cell receptor recognition of MenY CPS vaccine antigen and formation of a single antibody type; (b) B-cell recognition of MenW CPS vaccine antigen and formation of multiple antibody types. Colour markings on antibodies indicate the specific type of antibody.

3.4 Conclusions

The aim of this molecular modelling study was to investigate structural conformations of meningococcal group Y and W CPS RUs and to correlate these to the observed differences in immunogenic activity. The single difference between the two primary structures is the orientation of the hydroxyl group at C-4: equatorial for MenY and axial for MenW.

Metadynamics of the disaccharides and MD simulations of the 3RUs revealed clear differences in the conformational dynamics of the structures studied. Conformational differences between the RUs were revealed to arise primarily as a result of the orientation of the N26Glc and N26Gal linkages. The N26Gal linkages are more flexible and able to adopt multiple rotameric conformations whereas the N26Glc linkages adopt a single preferred rotamer.

Consequently, clustering analysis confirmed a single preferred conformer for the MenY 3RU, while the MenW 3RU has five significant cluster groups. Extended helical structures are adopted by the largest cluster family of both the MenY and MenW 3RUs. The additional cluster groups of MenW are more folded with a larger number of solvent inaccessible areas.

The results from the conformational analysis suggests a correlation between the CPS conformation and the observed differences in antibody cross-protection. This may be attributed to the distribution of the

antibodies formed by the MenY and MenW vaccine antigens.

The MenY vaccine generates antibodies of a single type corresponding to the single preferred dominant MenY CPS conformation. This antibody likely has strong avidity with the MenW CPS allowing for high levels of antibody cross-protection. In contrast, the MenW vaccine antigen will generate multiple antibodies and it is likely that only a certain percentage of these have strong avidity with the MenY CPS. As a result, MenW vaccine, in comparison, can only weakly cross-protect against MenY bacteria.

Chapter 4. Synthesis of the *S. sonnei* O-SP FucNAc4N residue

4.1 Overview

Shigella is one of the leading causes of dysentery worldwide and is caused by four known bacterial groups: *S. dysenteriae*, *S. flexneri*, *S. boydii*, and *S. sonnei*. A Swiss collaborator is developing a glycoconjugate vaccine against two important groups of Shigella bacteria: *S. flexneri* serotype 2a and *S. sonnei*. The O-SP of S. flexneri is composed of rhamnose, glucose and glucosamine, which are common sugars. In contrast, the *S. sonnei* RU comprises rare amino and amino-uronic acid sugars (FucNAc4N and AltNAcA) that provide particular challenges to glycoconjugate vaccine development.

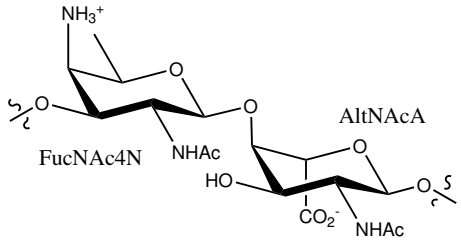


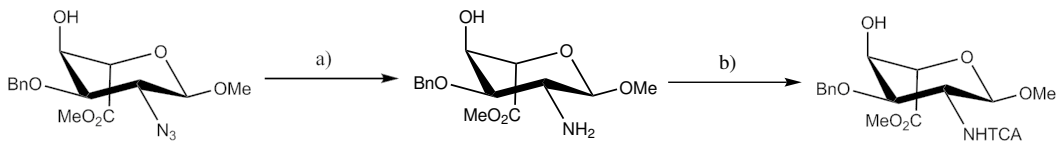
Figure 4.1: O-Specific polysaccharide repeating unit of *S. sonnei*

The development of a vaccine against shigellosis requires synthesis of the antigenic O-SP for biological testing and physicochemical characterisation. In this regard, our research group is developing a novel synthetic route toward the O-SP RU of *S. sonnei* (Figure 4.1).[147] Here, we synthesise a protected form of the FucNAc4N residue based on a synthetic route that is currently being researched in our laboratory group.

4.2 Synthetic routes toward the repeating unit of *S. sonnei*

4.2.1 Previous routes

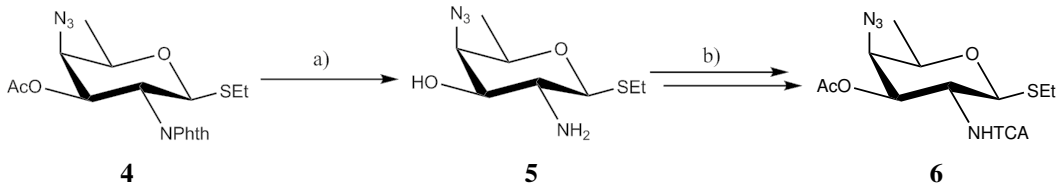
In 1999 Tòth *et al.* provided the first reported synthesis of the CPS repeating unit of *S. sonnei* in its zwitterionic form.[148] Scheme 4.1 shows the routes to the protected forms of AltNAcA as first reported by Tòth *et al*.



1 **2** **3**

Scheme 4.1: Synthesis of the protected form of AltNAc by Tòth et al.[148] Reagents and conditions: (a) H_2/Pd-C, MeOH, pyr, 25 °C, 2 h, 81%; (b) 1.1 equiv. of Cl_3CCOCl, Et_3N, CH_2Cl_2, 0 °C, 1 h, 70%

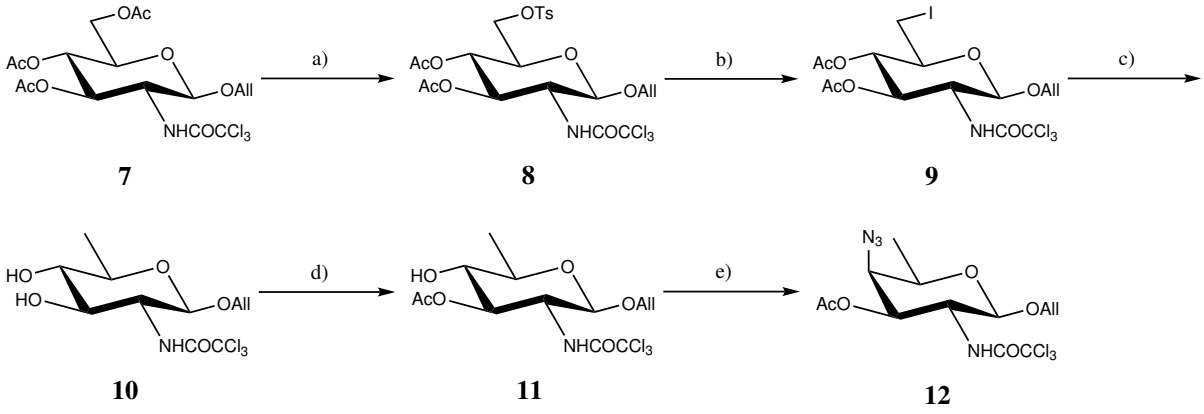
Synthesis of the protected form of AltNAcA, **3**, commenced from the L-glycoside (**1**) which was obtained from the commercially available but expensive L-glucose in 6 steps.[149] The method gave moderately good yields, although the greatest challenge lay in the synthesis of **1** from L-glucose, as the initial steps could only be undertaken at scales of approximately 1 g. This makes synthesising large scale quantities of AltNAc potentially difficult. In contrast, synthesis of the protected form of FucNAc4N, **6**, can be performed on significantly larger scales. Scheme 4.2 provides the route to the protected forms of FucNAc4N as first reported by Tòth *et al*.



4 **5** **6**

Scheme 4.2: Synthesis of the protected form of FucNAc4N by Tòth et al.[148] Reagents and conditions: (a) $NH_2CH_2CH_2NH_2$ (excess), EtOH, reflux, 2 h, 94%; (b) i. 1.2 equiv. of Cl_3CCOCl, Et_3N, CH_2Cl_2, 0 °C, 20 min, ii. pyr, Ac_2O, 25 °C, 2 h, 94%.

In 2014 Pfister *et al.* developed a novel route to the repeating units of *S. sonnei*.[150]. The synthesis of the AltNAcA residue (not shown here) is similar in method to that of the original route as it also commences from L-glucose. However, the final product was achieved in fewer steps and in moderately improved yields. The synthesis of the FucNAc4N residue commenced from **7** which was readily obtained from glucosamine hydrochloride. The subsequent steps involves selective reduction of C-6 by a sequence involving initial deacetylation of **7**, selective tosylation at O-6 and reacetylation to form **8**, displacement of the tosylate to form iodide **9**, and this was followed by preparation of a selectively protected derivative **11** by full deacetylation of di-acetate **9** followed by selective, low-temperature acetylation of O-3. Azide **12** was then prepared by initial formation of the 4-triflate from **11**, followed by an S_N2 of the azide (Scheme 4.3).



7 **8** **9**

10 **11** **12**

Scheme 4.3: Synthesis of the protected form of FucNAc4N by Pfister et al.[150] Reagents and conditions: (a) i. NaOMe, MOH, rt, 30 min, ii. TsCl, pyr, 0 °C, 3 h, iii. Ac_2O, rt, 3 h 88%; (b) NaI, Butanone, reflux, 4 h, 83%; c) i. $NaBH_3CN$, DMF, 110 °C, 30 h, ii. NaOMe, MeOH, rt, 3 h 86%; (d) AcCl, pyr, THF, -80 °C to rt, 16 h, 90%; (e) i. Tf_2O, pyr, DCM, 0 °C, 30 min, ii. NaN_3, DMF, rt, 2 h, 82%; (f) Na_2CO_3, MeOH, rt, 16 h, 92%.

In this study we aim to synthesise the O-SP RU of *S. sonnei* in good yields, using cheap, readily available starting materials.

4.2.2 Novel route towards FucNAc4N employing a 2, 3 oxazolidinone protected intermediate

A novel route to the D-FucNAc4N and L-AltNAcA units of *S. sonnei* has recently been under investigation by our research group. The study focuses on a divergent synthesis via the key 2,3-oxazolidinone protected glucosamine intermediate (Figure 4.2). Synthesis of the intermediate was first demonstrated by Benakli *et al* in 2001 in a simple and highly efficient route from N-acetyl glucosamine, which can be undertaken on a large scale.

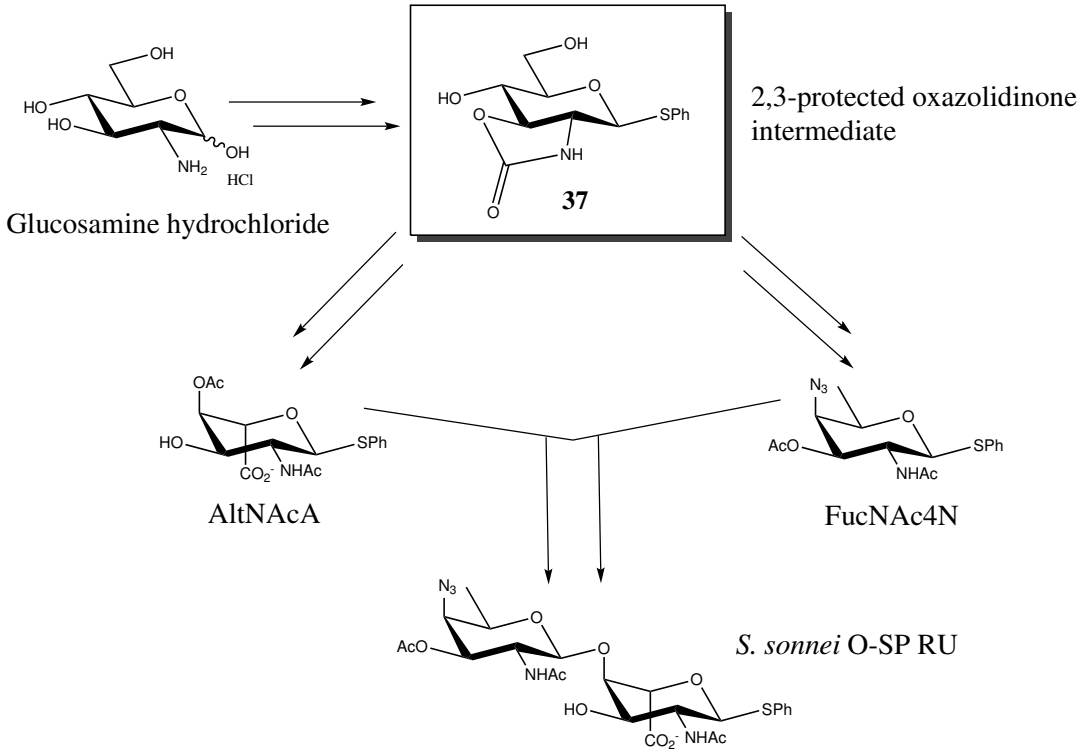
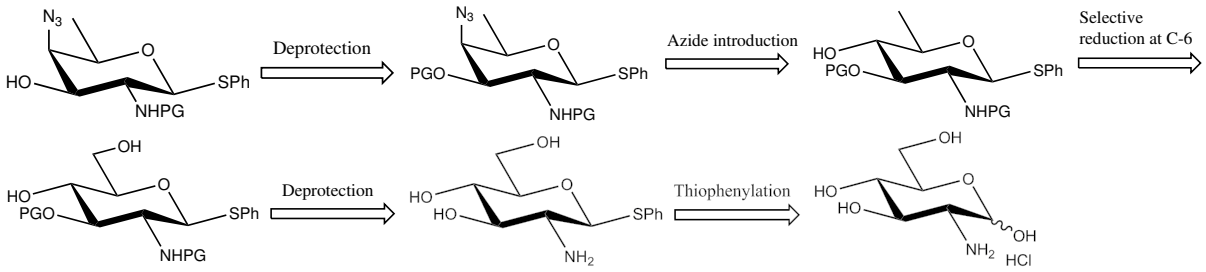


Figure 4.2: Proposed synthesis of the *S. sonnei* O-SP repeating unit from N-acetyl-D-glucosamine hydrochloride via the 2,3-protected oxazolidinone protected glucosamine intermediate.

An approach to the synthesis of D-FucNAc4N starting from D-glucosamine hydrochloride is illustrated in the retro-synthetic analysis provided in Scheme 4.4.



Scheme 4.4: Retrosynthetic analysis of the protected derivative of FucNAc4N from glucosamine hydrochloride

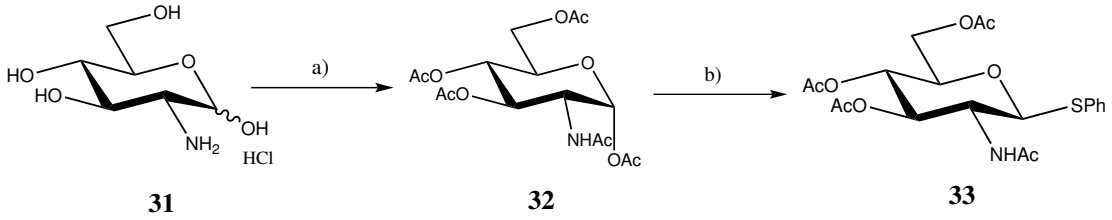
Formation of the axial azide at C-4, in the otherwise protected 4,6-dideoxy galactoside, requires substitution of the 4-OH with inversion in the 4,6-dideoxy glucoside. De-oxygenation at C-6 can be achieved

by selective primary iodination in the precursor diol, followed by reduction. The 4,6-diol in turn arises from selective protection of the 2-amino and 3-hydroxyl groups in phenyl 2-amino-1-thio-D-glucoside, which can be readily obtained by thiophenylation of N-actyl glucosamine. In this study we use the above approach and methodology developed by the group to synthesise the FucNAc4N residue of *S. sonnei.*.

4.3 Results and discussion

4.3.1 Synthesis of the β thioglycoside (33)

Synthesis of the 2, 3-oxazolidinone protected glucosamine intermediate **37**, commenced from the inexpensive and readily available glucosamine hydrochloride (**31**). The standard direct acetylation of this starting material provided N-Acetyl-D-glucosamine pentacetate (**32**) in high yields.



31 **32** **33**

Scheme 4.5: Synthesis of the β thioglycoside intermediate (**33**). Reagents and conditions: (a) Ac_2O, pyr, 25 °C, 4 h, 94%; (b) HSPh, $SnCl_4$, DCM, 60 °C, 20 h, 91%

Synthesis of the kinetically favoured α anomer of N-acetyl-D-glucosamine was confirmed by ^{1}H NMR. A single anomeric signal can be identified as a doublet at 6.14 ppm with a coupling constant of J = 3.67 Hz, confirming the sole presence of the α anomer. Acetylation of the amine and the hydroxyl groups was confirmed by the presence of five 3H singlets between 2.17 and 1.91 ppm. Assignment of the signals was done by comparison to literature values. NMR spectra for the structures synthesised are provided in Appendix B.

The second step involves conversion of the glycosyl acetate to a phenylthioglycoside. The presence of the N-Acetyl group at C-2 allows for the stereoselective formation of the β thiophenyl anomer (**33**). The reaction is Lewis acid catalysed by $SnCl_4$, which coordinates to the carbonyl of the anomeric acetate. The mechanism for the formation of the β thioglycoside is shown in Figure 4.3.

The ^{1}H NMR spectrum for **33** shows 4 methyl singlets between 1.95 and 2.05 ppm confirming the presence of the three OAc and one NAc groups. A triplet and doublet is found between 7.48 and 7.26 ppm for 5 protons indicating the successful addition of the thiophenyl group. The doublet at 4.83 ppm with J = 10.1 Hz confirms the sole formation of the β product.

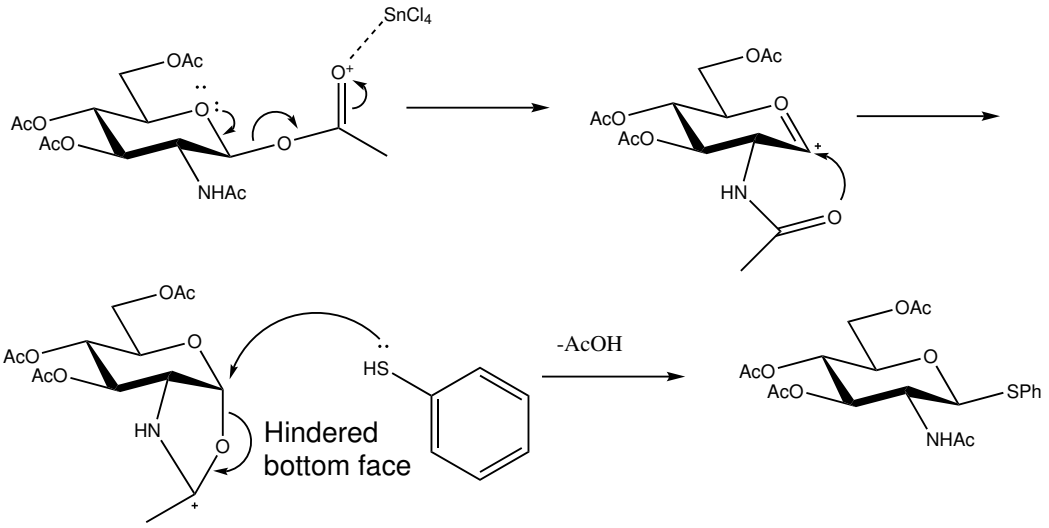
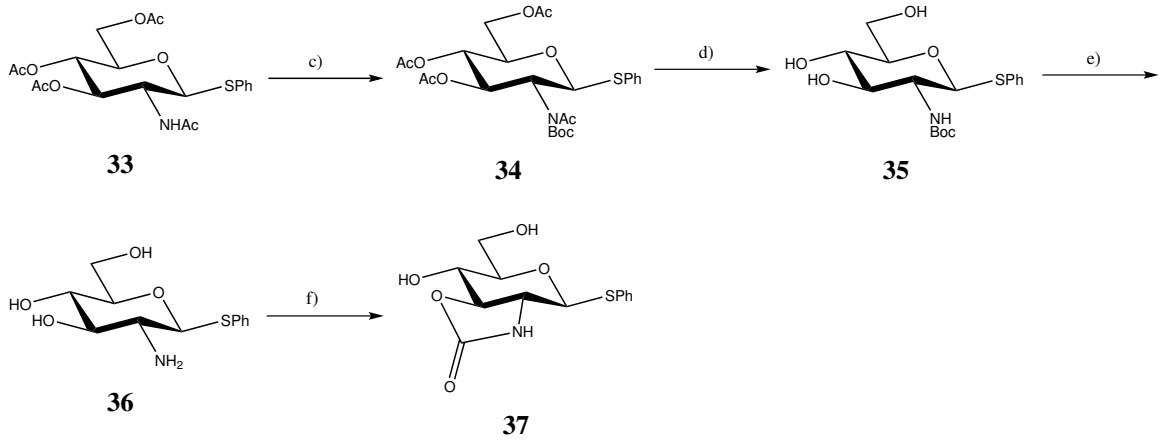


Figure 4.3: Mechanism for β-phenylthioglycoside formation.

4.3.2 Synthesis of the 2, 3-oxazolidinone protected glucosamine intermediate (37)

Synthesis of the 2, 3-oxazolidinone protected glucosamine intermediate, **37**, as described by Benakli *et al* requires a free amine and hydroxyl at C-2 and C-3 respectively (Scheme 4.6).[151] The first step is to form free hydroxyl groups at C-3, C-4, and C-6 employing the standard de-acetylation methodology of NaOMe in methanol. This method is sufficient to deprotect the hydroxyl groups but not the secondary amine. In order to achieve N-deacetylation a tert-butyloxycarbonyl (Boc)-protecting group is first added to the nitrogen under standard N-Boc protection methodology to form **34** in good yields.



Scheme 4.6: Synthesis of the 2, 3-oxazolidinone protected glucosamine intermediate (**37**). Reagents and conditions: (c) Boc_2O, DMAP, THF, 65 °C, 4 h, 92%; (e) NaOMe, MeOH, 25 °C, 2 h, 97%; (e) TFA, MeOH, 25 °C, 24 h, 71%; (f) 4-NPC, $NaHCO_3$, acetonitrile/H_20, 25 °C, 2.5 h, 83%.

The presence of the Boc group reduces the resonance into the amide carbonyl group with the result that the electrophilicity of the carbonyl carbon of the amide increases. Treatment with NaOMe is then able to deacetylate the N-Boc-amide efficiently to give **35** in high yields of over 95%. Synthesis of **34** was confirmed by NMR. The loss of the amide proton signal and the presence of the signal for the tertiary butyl protons confirmed the Boc protection of the amide. The formation of the deacetylated intermediate **35** was also confirmed by ^{1}H NMR by the absence of the series of acetyl singles on the hydroxyls and amide. The set of NMR spectra for these products can be found in Appendix B.

Subsequent deprotection of N-Boc to provide the free amine (**36**) can be achieved with a mild acid. It is reported that this can be performed with trifluoroacetic acid (TFA) in methanol to provide the product with very goods yields.[151] However, in our experience removal of the Boc group from the amide using this method was achieved in yields of less than 50%. Moreover, efficiency of the reaction reduced considerably with reactions on scale-up to several grams. Attempting the reaction in pure TFA at room temperature and increasing the reaction time from 16 h to 24 h increased the yield to approximately 65% - 70%. Nonetheless, the efficiency of the reaction remains highly dependent on scale with yields being significantly reduced with reactions of more than 2 grams.

Under certain circumstances Boc-deprotection of a secondary amine can be achieved by heating under acidic conditions. However, in this case heating can result in hydrolysis of the thioglycoside. Alternate methods were explored including methodology employing Tetra-n-butylammonium fluoride (TBAF) and heating under reflux for 72 h to deprotect the Boc group. The reaction can be left on for longer periods allowing it to go to completion without the risk of elimination of the thiophenyl group. Nonetheless, the work up for this reaction requires an acid wash to remove the excess TBAF and this poses a challenge as **36** is also water soluble, making purification of the product more challenging.

The mechanism for the formation of the 2, 3-oxazolidinone protected glucosamine intermediate, **37**, from β-thioglucosamine is shown in Figure 4.4. The reaction proceeds by nucleophilic attack on the carbonyl carbon of 4-nitrophenyl chloroformate by the free amine on the sugar to form a nitrophenyl carbamate, which is then attacked by O-3 with ultimate release of the stabilized p-nitrophenolate anion. This results in a five membered ring forming with an oxazolidinone group between C-2 and C-3.

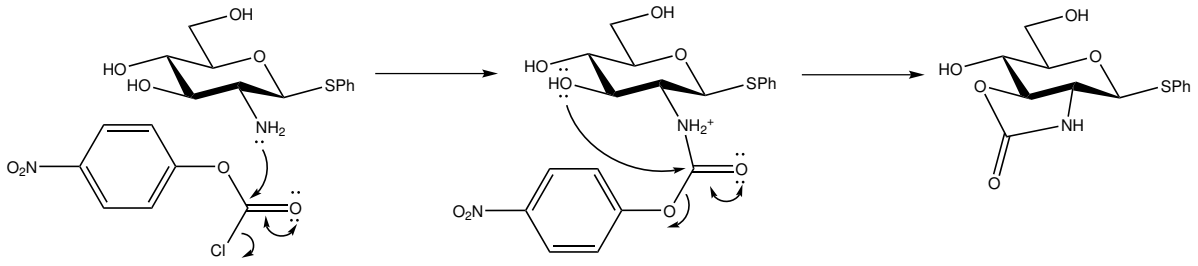


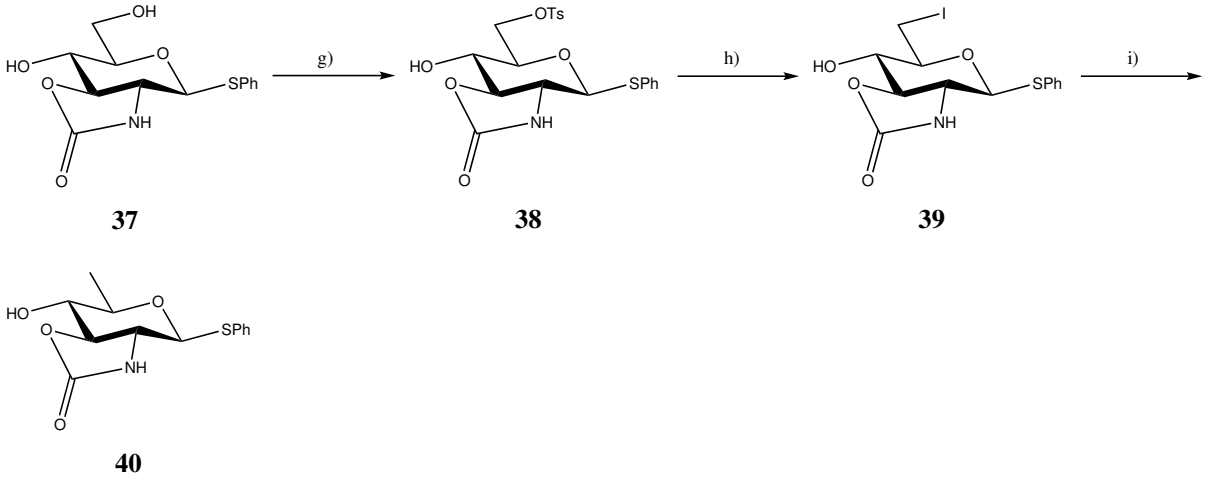
Figure 4.4: Mechanism for oxazolidinone intermediate formation

The reaction was undertaken on a scale of more than 5 grams and good yields (80%) of **26** were achieved. It was noted that if the 4-nitrophenyl choloformate is added at higher temperatures or too quickly multiple products form, as evident from spots appearing during TLC analysis that have similar R_f values and that are not readily separable by column chromatography. The formation of **37** was confirmed by

^{1}H NMR and where the assignment of protons was ambiguous comparisons were made to literature values.[151]

4.3.3 De-oxygenation at C-6 for the synthesis of Phenyl 2-amino-2-N,3-O-carbonyl-2,6-dideoxy-1-thio-β-D-glucopyranoside (40)

The next step in the synthesis of D-FucNAc4N involves the de-oxygenation at C-6, which can be achieved by selective primary tosylation of **37**, followed by iodination and reduction (Scheme 4.7). Synthesis of the tosylate (**38**) was performed at low temperature using one equivalent of tosylchloride to allow for selective tosylation solely at C-6. Product **38** was isolated in good yields.

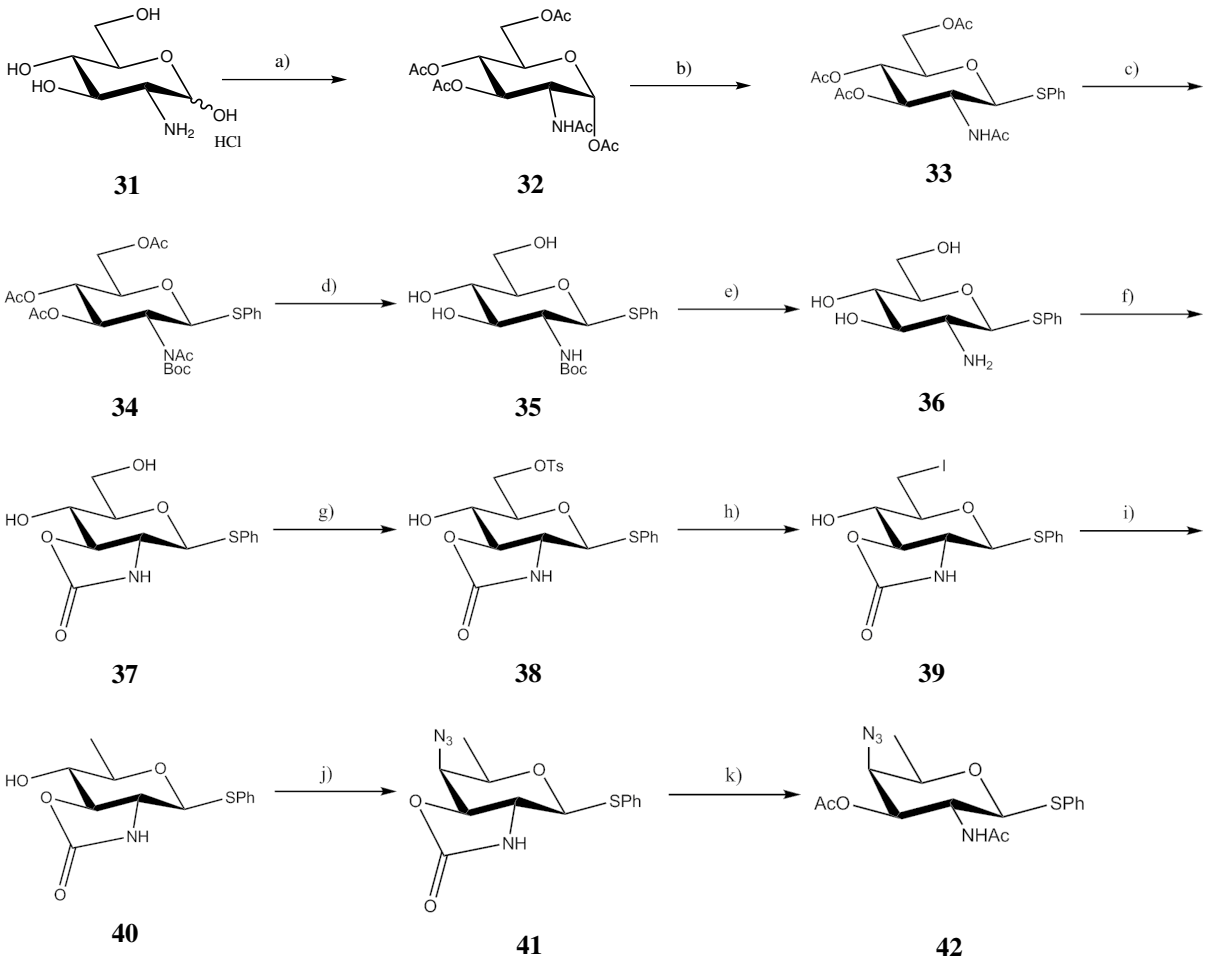


Scheme 4.7: Synthesis of Phenyl 2-amino-2-N,3-O-carbonyl-2,6-dideoxy-1-thio-β-D-glucopyranoside (**40**). Reagents and conditions: (g)TsCl, pyr, 0 °C, 2 h, 86%; (h) NaI, butanone, 25 °C, 6 h, 71%; (i) H_2/PdC 25 °C, 24 h, 67%.

A functional group interconversion was carried out on **38** to convert the tosyl- to an iodo- group. Subsequent halogenation under H_2/PdC conditions results in reductive elimination of the iodine and formation of a methyl group at C-6 to provide **40**. Synthesis of the product was confirmed by identifying signals of the five aromatic protons at 7.4 ppm as well as the presence of the C-6 methyl singlet near 1.4 ppm.

4.3.4 Synthesis of the protected form of FucNAc4N, 42

The final two steps in the synthesis of FucNAc4N involved the azidation with inversion at C-4 followed by deprotection at C-2 and C-3. The second last step is performed by the addition of trifluoromethanesulfonic anhydride to **40** in pyridine at -30 °C to provide azide **41**. The final deprotection is carried out with NaOH in THF at 60 °C immediately followed by N-Acetylation with acetic anhydride in pyridine. These steps were not performed due to time constraints, however, the full Scheme for the synthesis of FucNAc4N is provided in Scheme 4.8 based on additional work recently conducted in our group.



Scheme 4.8: Synthetic route toward the protected form of FucNAc4N, **42** (Note that the final sequence **40**→**40**→**42** was completed by another member of the research group). Reactants and conditions: (a) Ac_2O, pyr, 25 °C, 4 h; (b) HSPh, $SnCl_4$, DCM, 60 °C, 20 h; (c) Boc_2O, DMAP, THF, 65 °C, 4 h; (d) NaOMe, MeOH, 25 °C, 2 h; (e) TFA, MeOH, 25 °C, 24 h; 4-NPC, $NaHCO_3$, acetonitrile/H_20, 25 °C, 6 h; (g)TsCl, pyr, 0 °C, 2 h; (h) NaI, butanone, 25 °C, 6 h; (i) H_2/PdC 25 °C, 24 h; (j) Inversion and azide formation; (k) 2,3-deprotection and acetylation

4.4 Conclusions

The synthesis carried out was part of a broader research project working toward a novel synthesis of the O-SP repeating unit of *S. sonnei*. The aim of this study was to synthesise the protected form of the FucNAc4N residue, **42**, of *S. sonnei* via the 2,3-oxazolidinone protected intermediate (**37**). The 2,3-oxazolidinone intermediate was synthesised successfully from N-acetyl glucosamine via the thio-glycoside based on existing methodology. However, one notable challenge was identified in that the deprotection of the N-Boc group in phenyl 2-amino-2-deoxy-1-thio-β-D-glucopyranoside (**34**) posed a greater challenge and provided significantly lower yields than has been previously reported.

Due to time constraints the final step achieved was de-oxygenation of C-6 to provide phenyl 2-amino-2-N,3-O-carbonyl-2,6-dideoxy-1-thio-β-D-glucopyranoside (**40**). The final protected form of the Fuc-NAc4N residue, **42**, can be readily achieved from **40** by inversion and azidation at C-4 followed by

de-protection and acetylation at C-2 and C-3. In addition, we synthesised the α and β anomer derivatives of the 2,3-protected oxazolidinone intermediate as part of investigations into a divergent synthesis of the AltNAcA residue of *S. sonnei* O-SP.

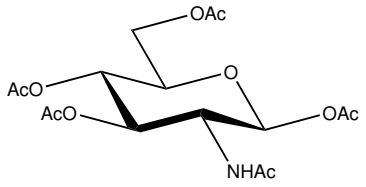
Chapter 5. Experimental

5.1 General procedures

All chemicals were purchased as reagent grade from Sigma-Aldrich. Unless otherwise noted, all solvents were purchased from Kimix (South Africa) as analytical reagent grade and used without further purification. Dichloromethane and toluene were distilled over calcium hydride. All reactions were monitored by analytical thin layer chromatography on silica gel 60 F_{254} pre-coated on aluminum plates. Spots were visualised by UV (254 nm) light and/or staining with ceric ammonium sulphate, followed by heating. Column chromatography was carried out using silica gel (200–300 mesh). ^{1}H NMR spectra were recorded with a Varian 300 MHz spectrometer at 30 °C. Chemical shifts (in ppm) were referenced to tetramethylsilane (δ=0 ppm) in deuterated chloroform. ^{13}C NMR spectra were recorded using the same NMR spectrometers and were calibrated with $CDCl_3$ (δ=77.00 ppm).

5.2 Synthesis

2-Acetamido-1,3,4,6-tetra-O-acetyl-α-D-glucopyranose (32)

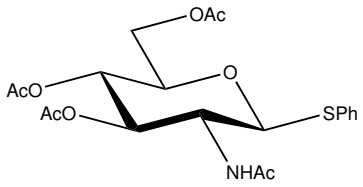


32

Acetic anhydride (26 ml, 0.28 mol) was added to a stirred solution of N-Acetyl glucosamine **31** (6.21 g, 28.8 mmol) in pyridine (20 mL) under argon. The reaction mixture was left to stir for 6 h at 25 °C when TLC showed the reaction had gone to completion. The resulting solution was diluted with 100 ml DCM and washed with 1 M HCl solution (3 x 100 ml). The organic layer was recovered and dried over sodium sulphate, filtered and concentrated under reduced pressure to provide **32** as a while solid (10.43 g, 93%). R_f = 0.29 (1:1 Ethyl acetate: Hexane); m.p. 184-186 °C (*lit.*[152] 186-189 °C).

^{1}H NMR (300 MHz, $CDCl_3$) δ ppm: 6.15 (d, *J*=3.67 Hz, 1H, H-1), 5.54 (d, *J*=8.80 Hz, 1H, H-3), 5.14 - 5.25 (m, 2H, H-4), 4.52 - 4.50 (m, 1H, H-6'), 4.23 (dd, *J*=4.03 , 12.54 Hz, 1H, H-6), 4.04 (dd, *J*=2.43, 12.49 Hz, 1H, H-5), 3.94 - 3.99 (m, 1H, H-2), 2.17, 2.06, 2.03, 2.02, 2.91 (5s, 15H, 5 x CH_3). ^{13}C NMR (100 MHz, $CDCl_3$) ppm: 171.72 (C=0), 170.65 (C=0), 169.90 (C=0), 169.05 (C=0), 168.55 (C=0), 90.17 (C-1), 77.43 (C-5), 77.01 (C-3), 76.58 (C-4), 61.56 (C-6), 51.10 (C-2), 23.02 (CH_3CO), 20.90 (CH_3CO), 20.69 (CH_3CO), 20.66 (CH_3CO), 20.53 (CH_3CO).

Phenyl 2-acetamido-3,4,6-tri-*O*-acetyl-2-deoxy-1-thio-β-D-glucopyranoside (33)

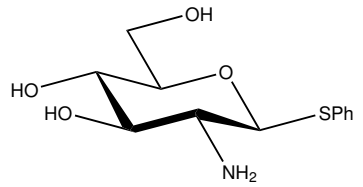


33

Thiophenol (3.5 ml, 31.7 mmol, 1.2) was added to a stirred solution of glycosyl acetate **32** (10.03 g, 25.8 mmol) and activated 4 Å molecular sieves (2 g) in DCM (50 ml) under Ar at 25 °C. To this, a solution of $SnCl_4$ (1 M in DCM, 3.5 ml, 25.7 mmol) was added. The reaction mixture was heated under reflux at 45 °C for 24 hrs after which the reaction mixture was allowed to cool to room temperature. The molecular sieves were filtered off using celite and quenched with the addition of $NaHCO_3$ (150 ml). The resulting mixture was washed with DCM (3 x 100 ml) and the organic layer was dried with sodium sulfate, then filtered and the solvent removed under vaccuum. The final product was recrystallised from hexane containing a minimum amount of DCM to produce **33** as a white solid (10.30 g, 91%). R_f = 0.27 (3:2 Ehthyl acetate: Hexane); m.p. 190-192 °C (*lit.*[152] 199 °C).

^{1}H NMR (300 MHz, $CDCl_3$) ppm: 7.48 – 7.42 (m, 2H, Ar-H), 7.29 – 7.26 (m, 3H, Ar-H), 5.58 (d, *J*=9.02 Hz, 1H, NH), 5.22 (dd, *J*=10.09, 9.48 Hz, 2H, H-3), 4.90 (dd, *J*=9.94, 9.48 Hz, 2H, H-4), 4.83 (d, *J*=10.4 Hz, 1H, H-1), 4.20-3.98 (m, 2H, H-6, H-6'), 4.06 – 3.98 (q, *J*=10.41, 9.17 Hz, 1H, H-2), 3.73 – 3.69 (m, 1H, H-5), 2.06, 2.01, 2.01, 1.97 (4s, 12H, 4 x CH3); ^{13}C NMR (100 MHz, $CDCl_3$) ppm: 171.0 (C=0), 169.5 (C=O), 169.3 (C=O), 169.1 (C=O), 132.8 (Ar), 131.4 (Ar), 128.8 x Ar), 85.9 (C-1), 79.2 (C-5), 76.5 (C-3), 74.2 (C-4), 61.3 (C-6), 55.2 (C-2), 21.3 (CH_3CO), 20.9 (CH_3CO), 20.8 (CH_3CO), 20.8 (CH_3CO), 20.7 (CH_3CO). NMR spectra agree with literature values.[151]

Phenyl 2-amino-2-deoxy-1-thio-β-D-glucopyranoside (36)



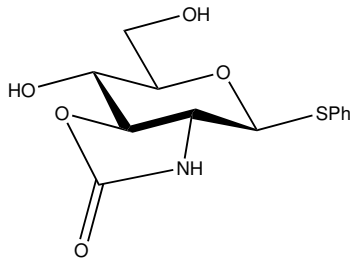
36

A catalytic amount of freshly prepared 1 M NaOMe solution was added to an ice cooled stirred solution of **34** (4.72 g, 13.7 mmol) in a minimum volume of MeOH. The reaction mixture was allowed to stir for 10 min, slowly brought to 25 °C and stirred for a further 60 min. Once TLC confirmed the starting material had fully reacted 5 g of Amberlyst IR-120 were added and the reaction mixture was allowed to stir for a further 20 min to provide **35**. The amberlyst was removed by filtration and TFA (100 ml, 1.31 mol, 100) was added to in 20 ml aliquots to over a 2.5 hour period and the reaction mixture was left to stir at 25 °C for a total of 24 hrs or until TLC showed the reaction went to completion. Excess TFA was

removed under reduced pressure and the solution was purified under a short column of silica gel (9.5:5 EtOAc:MeOH) to provide **36** as a brown oil which solidifies into a gum (1.69 g, 71%). Rf = 0.25 (9.5:5 EtOAc:MeOH)

^{1}H NMR (300 MHz, D_2O) ppm: 7.52 – 7.48 (m, 2H, Ar-H), 7.32 – 7.28 (m, 3H, Ar-H), 4.84 (d, *J*=10.4 Hz, 1H, H-1), 4.22 (dd, *J*=12.2, 5.3 Hz, 1H, H-6), 4.16 (dd, *J*=12.2, 2.6 Hz, 1H, H-6'), 4.03 (q, 1H, H-3), 3.75 – 3.66 (m, 2H, H-5, H-4), 3.49 (q, 1H, H-2); ^{13}C NMR (100 MHz, D_2O) ppm: 132.94 (Ar-C), 130.56 (Ar-C), 127.85 (2 x Ar-C), 125.69 (2 x Ar-C), 85.27 (C-1), 80.68 (C-5), 71.95 (C-3), 68.04 (C-4), 61.27 (C-6), 55.71 (C-2). NMR spectra agree with literature values.[147,151]

Phenyl 2-amino-2-N,3-O-carbonyl-2-deoxy-1-thio-β-D-glucopyranoside (37)



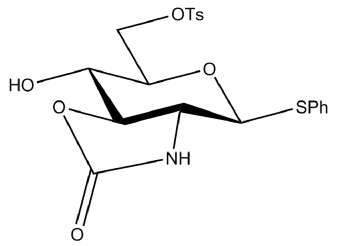
37

To a vigorously stirred, ice cooled solution of **36** (5.03 g, 16.9 mmol) in 200 ml water, and $NaHCO_3$ (7.11 g, 84.7 mmol), an ice cooled solution of *p*-nitrophenoxycarbonyl chloride (8.53 g, 42.3 mol) in acetonitrile (50 ml) was added drop-wise over 20 minutes. The reaction mixture was stirred for 2 hrs at 25 °C until TLC showed the reaction went to completion. The resulting solution was extracted with EtOAc (3x50 ml), dried over Na_2SO_4, concentrated, and purified using flash chromatography (3:7 Hexane:EtOAc to 1:9 Hexane:EtOAc) to provide 37 (4.57 g, 83%) as a yellow oil. R_f = 0.55 (EtOAc)

^{1}H NMR (300 MHz, CD_3OD) ppm: 7.68 – 7.65 (m, 2H, Ar-H), 7.43 – 7.39 (m, 3H, Ar-H), 4.99 (d, *J*=9.1 Hz, 1H, H-1), 4.25 – 4.13 (m, 1H, H-3), 3.97 - 3.79 (m, 3H, H-4, H-6, H-6'), 3.55 – 3.50 (m, 1H, H-5), 3.39 – 3.37 (m, 1H, H-2); ^{13}C NMR (100 MHz, CD_3OD) ppm: 162.11 (C=O), 133.73 (Ar), 133.57 (Ar), 130.24 (Ar), 129.31 (Ar), 86.27 (C-1), 86.21 (C-5), 84.04 (C-3), 68.67 (C-4), 62.23 (C-6), 61.70 (C-2). NMR spectra agree with literature values.[147,151]

Phenyl 2-amino-2-N,3-O-carbonyl-2-deoxy-1-thio-6-O-tosyl-β-D-glucopyranoside (38)

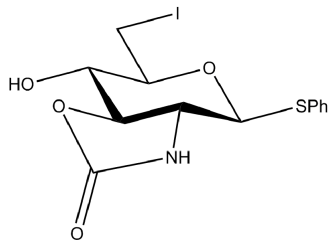
A solution of **37** (3.14 g, 10.5 mmol) and activated 4 Å molecular sieves was stirred in dry pyridine (25 ml) under argon and cooled to 0 °C. Freshly recrystallised *p*-toluenesulfonyl chloride (2.15 g, 12.7 mmol) was added to this and stirred for 6 hrs at 0 °C. The reaction mixture was allowed to warm to room temperature, filtered through celite, washed with 1 M HCl (3x20 ml) and the combined EtOAc extracts dried over Na_2SO_4. The reaction mixture was concentrated and purified using a short column of silica gel (1:1 EtOAc:Hexane) to yield **38** (4.10 g, 86%)as a yellow oil. R_f = 0.61 (EtOAc).



38

^{1}H NMR (300 MHz, $CDCl_3$) ppm: 7.47–7.43 (m, 2H, Ar-H), 7.32–7.28 (m, 3H, Ar-H), 7.47–7.43 (m, 5H, Ar-H), 5.08 (s, NH), 4.70 (d, *J*=9.6, 1H, H-1), 4.36 (dd, *J*=2.15, 1H, H-6), 4.37 (dd, *J*=2.16, 1H, H-6'), 4.10 (t, *J*=9.81, 1H, H-3), 3.98 (td, *J*=2.89, 1H, H-4), 3.56–3.6 (m, 1H, H-5), 3.26–3.33 (ddd, *J*=1.31, 1H, H-2), 2.44 (s, 3H, CH_3); ^{13}C NMR (100 MHz, $CDCl_3$) ppm: 158.49 (C=0), 145.46 (Ar), 133.82 (Ar), 132.88 (Ar), 130.23 (Ar), 130.13, (Ar), 129.53 (2xAr), 128.24 (Ar), 84.67 (C-1), 84.07 (C-3), 79.23 (C-5), 67.89 (C-6), 67.39 (C-4), 58.28 (C-2), 21.88 (CH_3). NMR spectra agree with literature values.[147]

Phenyl 2-amino-2-N,3-O-carbonyl-2,6-dideoxy-6-iodo-1-thio-β-D-glucopyranoside (39)

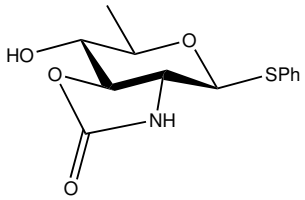


39

Dry butanone (17 ml) was added to a mixture of NaI (2.60 g, 6.37 mmol) and **38** (1.44 g, 3.19 mmol) under Argon. The reaction mixture was stirred at 80 °C for 10 hrs. The reaction mixture was concentrated and purified using flash chromatography to provide **39** (0.92 g, 71%) as a white solid. R_f = 0.68 (EtOAc); m.p. 161-164 °C (*lit.*[147] 167 °C-169 °C).

^{1}H NMR (300 MHz, $CO(CD_3)_2$) ppm: 7.68–7.65 (m, 2H, Ar-H), 7.35–7.32 (m, 3H, Ar-H), 4.75 (d, *J*=9.58 Hz, 1H, H-1), 4.59 (s, 1H, NH), 4.25–4.20 (m, 1H, H-3), 3.97–3.79 (m, 3H, H-4, H-6, H-6'), 3.5–3.55 (m, 1H, H-5), 3.29–3.45 (m, 1H, H-2); ^{13}C NMR (100 MHz, $CO(CD_3)_2$) ppm: 159.37 (C=O), 133.21 (Ar-C), 132.71 (Ar-C), 131.05 (2 x Ar-C), 129.26 (2 x Ar-C), 86.23 (C-1), 85.53 (C-3), 81.27 (C-5), 74.14 (C-4), 62.99 (C-2), 5.39 (C-6). NMR spectra agree with literature values.[147]

Phenyl 2-amino-2-N,3-O-carbonyl-2,6-dideoxy-1-thio-β-D-glucopyranoside (40)



40

39 (1.47 g, 3.61 mmol) and $NaHCO_3$ were dissolved in MeOH (6 ml) followed by the addition of a mixture of 10 wt % Pd/C (1.04 g) in MeOH. The reaction mixture was flushed with H_2 allowed to remain in a hydrogenator for 24 hrs at 4 bar. The reaction mixture was subsequently filtered through celite and purified using flash chromatography to provide **40** (2.68 g, 67%) as a white solid. R_f = 0.64 (7:3 EtOAc:Hexane); m.p. 167-170 °C. (*lit.*[147])

^{1}H NMR (300 MHz, $CDCl_3$) ppm: 7.59–7.55 (m, 2H, Ar-H), 7.43–7.40 (m, 3H, Ar-H), 4.99 (d, *J*=9.58, 1H, H-1), 4.17–4.25 (m, 1H, H-3), 3.57–3.61 (m, 1H, H-4), 3.43–3.47 (m, 1H, H-5), 3.11–3.14 (m, 1H, H-2), 2.52 (d, 1H, OH), 1.41 (d, *J*=5.8, 3H, H-6); ^{13}C NMR (100 MHz, $CDCl_3$) ppm: 158.93 (C=O), 133.38 (Ar-C), 131.05 (Ar-C), 130.01 (2 x Ar-C), 128.99 (2 x Ar-C), 86.95 (C-3), 86.67 (C-1), 77.99 (C-5), 72.32 (C-4), 59.83 (C-2), 17.08 (C-6). NMR spectra agree with literature values.[147]

Chapter 6. Conclusions

This study has two main components. The aim of the molecular modelling analysis was to investigate structural conformations of meningococcal group Y and W CPS RUs and to correlate these to observed differences in immunogenic activity. The sole difference between the two primary structures is the orientation of the hydroxyl group at C-4: equatorial for MenY and axial for MenW.

The first phase in the computational modelling involved solution metadynamic simulations of four disaccharides: Glc14N, Gal14N, N16Glc, and N26Gal. The single preferred conformation of the Glc14N and Gal14N disaccharides are almost identical. The dominant conformations of N26Glc and N26Gal are also similar, although, the results suggest that the N26Gal linkage can rotate more freely than the N26Glc linkage.

The second phase comprised of MD simulations of MenY and MenW 3RUs. Our calculations showed clear differences in the solution conformations of the two structures. The molecular conformations of the two RUs correspond to the orientation of the N26Glc and N26Gal linkages resulting in the MenY 3RU having less conformational flexibility than the MenW 3RU.

Clustering analysis of the structures from the MD trajectories confirm a single preferred conformer for the MenY 3RU, with over 95% of the structures populating this group. The MenW 3RU has five significant cluster groups. The largest group is conformationally similar to that of the MenY 3RU cluster and is formed by 44% of the structures. The remaining four clusters of MenW have significantly smaller populations.

The structural analysis indicates a correlation between CPS conformation and the observed differences in cross-protection of MenY and MenW CPS vaccines. The results suggest that the MenY vaccine may be able to generate a sufficiently large number of antibodies with high avidities, targeting the dominant CPS conformations of the MenW bacterium and allowing for strong cross-protection. However, the MenW vaccine produces a range of multiple antibodies but a smaller amount of the antibodies against the single preferred CPS conformation of the MenW bacterium, resulting in weaker levels of cross-protection.

We conducted the conformational analysis on the dOA RUs of MenY and MenW. However, the CPS of the two meningococcal groups may be partially O-acetylated at C-7 and C-9 of the sialic acid residue. The structural nature of MenY with respect to different levels of O-acetylation, has been previously investigated spectroscopically.[153,154] The effect of O-acetylation on molecular conformation can be further studied for both MenY and MenW CPS RUs particular as there is a known correlation between O-acetylation and immunogenic activity.[154]

The experimental component was part of the broader research project to develop a novel synthetic route the the O-SP of *S. sonnei*. We have successfully developed a novel route toward the FucNAc4N residue of *S. sonnei* from glucosamine hydrochloride via a 2, 3-oxazolidinone protected glucosamine intermediate. Additional studies we also performed on the 2,3-oxazolidinone intermediate as part of a divergent synthesis strategy toward the AltNAc residue of *S. sonnei*.

If required we will continue to explore alternate methods[155] toward the 2,3-oxazolidinone intermediate that may provide improved yields. As part of the on going research we will also work the divergent synthesis methodology to achieve both the FucNAc4N and AltNAcA residues from a common starting material.

Bibliography

1. L. Rothfield, *Struct. Funct. Biol. Membr.*, 1971, 3–9.
2. T. J. Silhavy, D. Kahne and S. Walker, *Cold Spring Harb. Perspect. Biol.*, 2010, **2**, a000414.
3. C. Jones, *An. Acad. Bras. Cienc.*, 2005, **77**, 293–324.
4. N. E. Rosenstein, B. A. Perkins, D. S. Stephens, T. Popovic and J. M. Hughes, *N. Engl. J. Med.*, 2001, **344**, 1378–1388.
5. C. Whitfield, *Can. J. Microbiol.*, 1988, **34**, 415–420.
6. L. Deng, D. L. Kasper, T. P. Krick and M. R. Wessels, *J Biol Chem*, 2000, **275**, 7497–7504.
7. C. Whitfield and M. A. Valvano, *Adv. Microb. Physiol.*, 1993, **35**, 135–264.
8. L. M. Willis and C. Whitfield, *Carbohydr. Res.*, 2013, **378**, 35–44.
9. I. S. Roberts, *Microbiology*, 1995, **141**, 2023–2031.
10. S. Pelkonen, J. Häyrinen and J. Finne, *J. Bacteriol.*, 1988, **170**, 2646–2653.
11. K. D. Kröncke, J. R. Golecki, K. Jann and K.-D. Kroncke, *J. Bacteriol.*, 1990, **172**, 3469–3472.
12. K. Bazaka, R. J. Crawford, E. L. Nazarenko and E. P. Ivanova, *Adv. Exp. Med. Biol.*, 2011, **715**, 213–226.
13. J. W. Costerton, R. T. Irvin and K. J. Cheng, *Annu. Rev. Microbiol.*, 1981, **35**, 299–324.
14. Y. Kang, S. Barbirz, R. Lipowsky and M. Santer, *J. Phys. Chem. B*, 2014, **118**, 2523–2534.
15. T. Galochkina, D. Zlenko, A. Nesterenko, I. Kovalenko, M. Strakhovskaya, A. Averyanov and A. Rubin, *ChemPhysChem*, 2016, 2839–2853.
16. M. M. Kuttel, G. E. Jackson, M. Mafata and N. Ravenscroft, *Carbohydr. Res.*, 2015, **406**, 27–33.
17. E. R. Moxon and J. S. Kroll, *Curr. Top. Microbiol. Immunol.*, 1990, **150**, 65–85.
18. C. A. Doige and G. F. Ames, *Annu. Rev. Microbiol.*, 1993, **47**, 291–319.
19. M. A. Horwitz and S. C. Silverstein, *J. Clin. Invest.*, 1980, **65**, 82–94.
20. P. J. Delves, S. J. Martin, D. R. Burton and I. M. Roitt, *Roitt's essential immunology*, Wiley-Blackwell, 12th edn, 2011.
21. F. A. Wyle, M. S. Artenstein, B. L. Brandt, E. C. Tramont, D. L. Kasper, P. L. Altieri, S. L. Berman and J. P. Lowenthal, *J. Infect. Dis.*, 1972, **126**, 514–522.
22. J. Charles A Janeway, P. Travers, W. Mark and M. J. Shlomchik, *Immunobiology*, Garland Science, New York, NY, 5th edn, 2001.
23. M.-I. Yuseff, P. Pierobon, A. Reversat and A.-M. Lennon-Duménil, *Nat. Rev. Immunol.*, 2013, **13**, 475–486.
24. A. Coutinho and G. Möller, *Nat. New Biol.*, 1973, **245**, 12–14.
25. A. Coutinho, G. Möller, J. Andersson and W. W. Bullock, *Eur. J. Immunol.*, 1973, **3**, 299–306.
26. J. R. McGhee, S. M. Michalek, H. Kiyono, J. H. Eldridge, D. E. Colwell, S. I. Williamson, M. J. Wannemuehler, E. Jirillo, L. M. Mosteller, D. M. Spalding, S. Hamada, K. A. Gollahon, I. Morisaki, R. L. Gregory and W. J. Koopman, *Microbiol. Immunol.*, 1984, **28**, 261–280.
27. D. J. Barrett, *Adv. Pediatr.*, 1985, **32**, 139–58.
28. G. Lesinski and M. Julie Westerink, *Curr. Drug Target -Infectious Disord.*, 2001, **1**, 325–334.

29. F. Y. Avci and D. L. Kasper, *Annu. Rev. Immunol.*, 2010, **28**, 107–130.
30. C.-J. Lee, L. H. Lee and K. Koizumi, *Infect. Med.*, 2002, **19**, 127–133.
31. B. A. Cobb, Q. Wang, A. O. Tzianabos and D. L. Kasper, *Cell*, 2004, **117**, 677–687.
32. J. Volpe, *Host Defenses II*, `http://spot.pcc.edu/{~}jvolpe/b/bi234/lec/8{_}9defenses/defensesII/fig16.12{_}T-independentBind.jpg`.
33. B. A. Cobb and D. L. Kasper, *Eur. J. Immunol.*, 2005, **35**, 352–356.
34. C. C. A. M. Peeters, P. R. Lagerman, O. de Weers, L. A. Oomen, P. Hoogerhout, M. Beurret, J. T. Poolman and K. M. Reddin, *Vaccine Protoc.*, Humana Press, New Jersey, 2003, pp. 153–174.
35. F. Y. Avci, X. Li, M. Tsuji and D. L. Kasper, *Semin. Immunol.*, 2013, **25**, 146–151.
36. T. I. Prigozy, O. Naidenko, P. Qasba, D. Elewaut, L. Brossay, A. Khurana, T. Natori, Y. Koezuka, A. Kulkarni and M. Kronenberg, *Science*, 2001, **291**, 664–667.
37. M. Heidelberger and O. T. Avery, *J. Exp. Med.*, 1923, **38**, 73–79.
38. M. Heidelberger and O. T. Avery, *J. Exp. Med.*, 1924, **40**, 301–317.
39. R. Austrian, *J Infect Dis*, 1975, **131**, 474–484.
40. P. Kazanjian, *J. Hist. Med. Allied Sci.*, 2004, **59**, 555–587.
41. J. D. Grabenstein and K. P. Klugman, *Clin. Microbiol. Infect.*, 2012, **18 Suppl 5**, 15–24.
42. J. Hütter and B. Lepenies, *Carbohydrate-Based Vaccines An Overview*, Springer New York, New York, NY, 2015, vol. 1331, pp. 1–10.
43. J. F. G. Vliegenthart, *FEBS Lett.*, 2006, **580**, 2945–2950.
44. R. D. Astronomo and D. R. Burton, *Nat. Rev. Drug Discov.*, 2010, **9**, 308–324.
45. WHO, *Weekly Epidemiological Record*, 2007, **82**, 93–104.
46. WHO, *Weekly Epidemiological Record*, 2008, **83**, 373–384.
47. WHO, *The Immunological Basis for Immunization Series - Module 9: Haemophilus influenzae type b*, 2007, `http://whqlibdoc.who.int/publications/2009/9789241597555{_}eng.pdf`.
48. WHO, *Weekly Epidemiological Record*, 2013, **88**, 413–428.
49. C. J. Lee, L. H. Lee and K. Koizumi, *Infect. Med.*, 2002, **19**, 179–182.
50. J. Ihssen, M. Kowarik, S. Dilettoso, C. Tanner, M. Wacker and L. Thöny-Meyer, *Microb. Cell Fact.*, 2010, **9**, 61.
51. A. J. Pollard, K. P. Perrett and P. C. Beverley, *Nat. Rev. Immunol.*, 2009, **9**, 213–220.
52. E. A. Kabat and A. E. Bezer, *Arch. Biochem. Biophys.*, 1958, **78**, 306–318.
53. E. C. Gotschlich, M. Rey, J. Etienne, W. R. Sanborn, R. Triau and B. Cvjetanović, *Prog. Immunobiol. Stand.*, 1971, **5**, 485–91.
54. a. a. Lindberg, L. T. Rosenberg, a. Ljunggren, P. J. Garegg, S. Svensson and N. H. Wallin, *Infect. Immun.*, 1974, **10**, 541–5.
55. S. C. Szu, X. R. Li, R. Schneerson, J. H. Vickers, D. Bryla and J. B. Robbins, *Infect. Immun.*, 1989, **57**, 3823–7.
56. L. Chijen, *Mol. Immunol.*, 1987, **24**, 1005–1019.
57. S. L. Moore, C. Uitz, C.-C. Ling, D. R. Bundle, P. C. Fusco and F. Michon, *Clin. Vaccine Immunol.*, 2007, **14**, 1311–7.
58. CDC, *Principles of Vaccination*, 2015, `http://www.cdc.gov/vaccines/pubs/`

pinkbook/prinvac.html.

59. D. L. Kasper, L. C. Paoletti, M. R. Wessels, H. K. Guttormsen, V. J. Carey, H. J. Jennings and C. J. Baker, *J. Clin. Invest.*, 1996, **98**, 2308–14.
60. M. E. Pichichero, S. Porcelli, J. Treanor and P. Anderson, *Vaccine*, 1998, **16**, 83–91.
61. M. D. Decker, K. M. Edwards, R. Bradley and P. Palmer, *J. Pediatr.*, 1992, **120**, 184–189.
62. WHO and Others, *Managing meningitis epidemics in Africa*, 2015, http://www.who.int/csr/resources/publications/HSE{_}GAR{_}ERI{_}2010{_}4/en/.
63. R. R. Carpenter and R. G. Petersdorf, *Am. J. Med.*, 1962, **33**, 262–275.
64. M. C. Brouwer, A. R. Tunkel and D. van de Beek, *Clin. Microbiol. Rev.*, 2010, **23**, 467–492.
65. WHO, *Global Vaccine Action Plan*, 2011.
66. M. P. Girard, M. P. Preziosi, M. T. Aguado and M. P. Kieny, *Vaccine*, 2006, **24**, 4692–4700.
67. D. S. Stephens, B. Greenwood and P. Brandtzaeg, *Lancet*, 2007, **369**, 2196–2210.
68. R. Hilse, S. Hammerschmidt, W. Bautsch and M. Frosch, *J. Bacteriol.*, 1996, **178**, 2527–32.
69. N. J. Raymond, M. Reeves, G. Ajello, W. Baughman, L. L. Gheesling, G. M. Carlone, J. D. Wenger and D. S. Stephens, *J. Infect. Dis.*, 1997, **176**, 1277–1284.
70. Y.-L. Tzeng and D. S. Stephens, *Microbes Infect.*, 2000, **2**, 687–700.
71. L. K. K. Tan, G. M. Carlone and R. Borrow, *N. Engl. J. Med.*, 2010, **362**, 1511–1520.
72. D. S. Stephens, *FEMS Microbiol. Rev.*, 2007, **31**, 3–14.
73. V. S. Joshi, I. B. Bajaj, S. A. Survase, R. S. Singhal and J. F. Kennedy, *Carbohydr. Polym.*, 2009, **75**, 553–565.
74. M. C. Schneider, R. M. Exley, S. Ram, R. B. Sim and C. M. Tang, *Trends Microbiol.*, 2007, **15**, 233–240.
75. L. A. Riley and B. P. Currie, *Clin. Infect. Dis.*, 2005, **41**, 277–278.
76. E. C. Gotschlich, I. GoldSchneider and M. S. Artenstein, *J. Exp. Med.*, 1969, **129**, 1367–1384.
77. N. E. Rosenstein, B. A. Perkins, D. S. Stephens, L. Lefkowitz, M. L. Cartter, R. Danila, P. Cieslak, K. A. Shutt, T. Popovic, A. Schuchat, L. H. Harrison and A. L. Reingold, *J. Infect. Dis.*, 1999, **180**, 1894–1901.
78. J. S. Swartley, A. A. Marfin, S. Edupuganti, L. J. Liu, P. Cieslak, B. Perkins, J. D. Wenger and D. S. Stephens, *Proc. Natl. Acad. Sci. U. S. A.*, 1997, **94**, 271–276.
79. K. H. Dyet and D. R. Martin, *Epidemiol. Infect.*, 2006, **134**, 377–383.
80. A. Wilder-smith, K. T. Goh, T. Barkham and N. I. Paton, *Clin. Infect. Dis.*, 2003, **36**, 679–683.
81. J. Decosas and J. B. T. Koama, *Lancet Infect. Dis.*, 2002, **2**, 763–765.
82. P. L. Raghunathan, J. D. Jones, S. R. M. Tiendrebe, I. Sanou, L. Sangare, S. Kouanda, M. Dabal, C. Lingani, C. M. Elie, S. Johnson, M. Ari, J. Martinez, J. Chatt, K. Sidibe, S. Schmink, L. W. Mayer, M. K. Konde, M. H. Djingarey, T. Popovic, B. D. Plikaytis, G. M. Carlone, N. Rosenstein and M. Soriano-Gabarró, *J. Infect. Dis.*, 2006, **193**, 607–16.
83. WHO, *The emergence of Neisseria meningitidis W135*, 2015, www.who.int/csr/disease/meningococcal/w135/en/.
84. S. N. Ladhani, K. Beebeejaun, J. Lucidarme, H. Campbell, S. Gray, E. Kaczmarski, M. E. Ramsay and R. Borrow, *Clin. Infect. Dis.*, 2015, **60**, 578–585.
85. Public Health England, *Continuing increase in meningococcal group W (MenW) disease in Eng-*

land, Public Health England Technical Report 7, 2015.

86. A. M. Molesworth, L. E. Cuevas, S. J. Connor, A. P. Morse and M. C. Thomson, *Emerg. Infect. Dis.*, 2003, **9**, 1287–1293.
87. H. Peltola, A. Safary, H. Käyhty, V. Karanko and F. E. André, *Pediatrics*, 1985, **76**, 91–96.
88. I. D. Committee, *Prevention and Control of Meningococcal Disease: Recommendations for Use of Meningococcal Vaccines in Pediatric Patients*, American Academy of Pediatrics Technical Report 2, 2005.
89. J. M. Griffiss, B. L. Brandt, P. L. Altieri, G. B. Pier and S. L. Berman, *Infect. Immun.*, 1981, **34**, 725–32.
90. K. L. Kotloff, J. P. Nataro, W. C. Blackwelder, D. Nasrin, T. H. Farag, S. Panchalingam, Y. Wu, S. O. Sow, D. Sur, R. F. Breiman, A. S. Faruque, A. K. Zaidi, D. Saha, P. L. Alonso, B. Tamboura, D. Sanogo, U. Onwuchekwa, B. Manna, T. Ramamurthy, S. Kanungo, J. B. Ochieng, R. Omore, J. O. Oundo, A. Hossain, S. K. Das, S. Ahmed, S. Qureshi, F. Quadri, R. A. Adegbola, M. Antonio, M. J. Hossain, A. Akinsola, I. Mandomando, T. Nhampossa, S. Acácio, K. Biswas, C. E. O'Reilly, E. D. Mintz, L. Y. Berkeley, K. Muhsen, H. Sommerfelt, R. M. Robins-Browne and M. M. Levine, *Lancet*, 2013, **382**, 209–222.
91. E. Yabuuchi, *Int. J. Syst. Evol. Microbiol.*, 2002, **52**, 1041.
92. S. K. Niyogi, *J. Microbiol.*, 2005, **43**, 133–143.
93. WHO, *Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1*, 2005.
94. F. Qu, C. Bao, S. Chen, E. Cui, T. Guo, H. Wang, J. Zhang, H. Wang, Y.-W. Tang and Y. Mao, *Diagn. Microbiol. Infect. Dis.*, 2012, **74**, 166–70.
95. E. M. Barry, M. F. Pasetti, M. B. Sztein, A. Fasano, K. L. Kotloff and M. M. Levine, *Nat. Rev. Gastroenterol. Hepatol.*, 2013, **10**, 245–255.
96. P. Cossart and P. J. Sansonetti, *Science*, 2004, **304**, 242–248.
97. A. Phalipon and P. J. Sansonetti, *Immunol. Cell Biol.*, 2007, **85**, 119–129.
98. M. M. Levine, K. L. Kotloff, E. M. Barry, M. F. Pasetti and M. B. Sztein, *Nat. Rev. Microbiol.*, 2007, **5**, 540–553.
99. M. Ward, *Sexual Health – Shigella*, 2014, `http://www.haemosexual.com/sexual-health-shigella/`.
100. CDC, *Foodborne Dis. Act. Surveill. Netw. FoodNet Surveill. Rep. 2011 (Final Report)*, 2012, 1–52.
101. K. C. Hyams, A. L. Bourgeois, B. R. Merrell, P. Rozmajzl, J. Escamilla, S. A. Thornton, G. M. Wasserman, A. Burke, P. Echeverria, K. Y. Green and E. al., *N Engl J Med*, 1991, **325**, 1423–1428.
102. L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen and E. Romanowska, *Eur. J. Biochem.*, 1977, **76**, 327–330.
103. L. Kenne, B. Lindberg, K. Petersson, E. Katznellenbogen and E. Romanowska, *Eur. J. Biochem.*, 1978, **284**, 279–284.
104. L. Kenne, B. Lindberg, K. Petersson, E. Katzenellenbogen and E. Romanowska, *Carbohydr. Res.*, 1980, **78**, 119–126.
105. J. Tao, L. Feng, H. Guo, Y. Li and L. Wang, *FEMS Microbiol. Lett.*, 2004, **234**, 125–132.
106. D. Cohen, M. S. Green, C. Block, R. Slepon and I. Ofek, *J. Clin. Microbiol.*, 1991, **29**, 386–389.

107. O. Guvench, E. Hatcher, R. M. Venable, R. W. Pastor and A. D. MacKerell, *J. Chem. Theory Comput.*, 2009, **5**, 2353–2370.
108. A. D. MacKerell, E. P. Raman, O. Guvench, A. D. MacKerell, E. P. Raman and O. Guvench, *J. Phys. Chem. B*, 2010, **114**, 12981–12994.
109. O. Guvench, S. S. Mallajosyula, E. P. Raman, E. Hatcher, K. Vanommeslaeghe, T. J. Foster, F. W. Jamison and A. D. MacKerell, *J. Chem. Theory Comput.*, 2011, **7**, 3162–3180.
110. S. S. Mallajosyula, O. Guvench, E. Hatcher and A. D. MacKerell, *J. Chem. Theory Comput.*, 2012, **8**, 759–776.
111. A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
112. B. Hess, H. Bekker, H. J. C. Berendsen and J. G. E. M. Fraaije, *J. Comput. Chem.*, 1997, **18**, 1463–1472.
113. W. F. Van Gunsteren and H. J. C. Berendsen, *Mol. Phys.*, 1977, **34**, 1311–1327.
114. W. C. Swope, *J. Chem. Phys.*, 1982, **76**, 637.
115. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *J. Chem. Phys.*, 1983, **79**, 926.
116. L. Zhigilei, *Boundary Conditions*, 2015.
117. *Metadynamics*, 2015, `http://www.ks.uiuc.edu/Research/vmd/current/ug/node233.html`.
118. L. J. Heyer, S. Kruglyak and S. Yooseph, *Genome Res.*, 1999, **9**, 1106–1115.
119. M. Kuttel, J. W. Brady and K. J. Naidoo, *J. Comput. Chem.*, 2002, **23**, 1236–1243.
120. M. Kuttel and K. J. Naidoo, *J. Am. Chem. Soc.*, 2005, **127**, 12–13.
121. J. A. Hadden, M. B. Tessier, E. Fadda and R. J. Woods, *Glycoinformatics*, Springer, 2015, pp. 431–465.
122. X. Lemercinier and C. Jones, *Carbohydr. Res.*, 1996, **296**, 83–96.
123. M. Kuttel, N. Ravenscroft, M. Foschiatti, P. Cescutti and R. Rizzo, *Carbohydr. Res.*, 2012, **350**, 40–48.
124. R. Stuike-Prill and B. Meyer, *Eur. J. Biochem.*, 1990, **194**, 903–913.
125. M. Kuttel, M. Gordon and N. Ravenscroft, *Carbohydr. Res.*, 2014, **390**, 20–27.
126. Y. Nishida, H. Ohrui and H. Meguro, *Tetrahedron Lett.*, 1984, **25**, 1575–1578.
127. Y. Nishida, H. Hori, H. Ohrui and H. Meguro, *J. Carbohydr. Chem.*, 1988, **7**, 239–250.
128. K. Bock and J. Duus, *Carbohydr. Chem.*, 1994, **13**, 513–543.
129. R. Stenutz, *The conformation of the hydroxymethyl group*, 2013.
130. O. Hassel, B. Ottar, B. Roald, A. Linnasalmi and P. Laukkanen, *The Structure of Molecules Containing Cyclohexane or Pyranose Rings*, 1947.
131. M. Kuttel, Y. Mao, G. G. G. Widmalm and M. Lundborg, *Proc. - 2011 7th IEEE Int. Conf. eScience, eScience 2011*, 2011, 395–402.
132. M. M. Kuttel, J. Ståhle and G. Widmalm, *J. Comput. Chem.*, 2016, 1–12.

133. J. Gullingsrud, J. Saam and J. Phillips, *PSFGEN User Guide*, 2016.
134. J. E. Stone, *VMD psfgen plugin*, http://www.ks.uiuc.edu/Research/vmd/plugins/psfgen/.
135. J. C. Phillips, R. Braun, W. E. I. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kal, K. Schulten and H. Poincare, *J. Comput. Chem.*, 2005, **26**, 1781–1802.
136. M. M. Kuttel, *PhD thesis*, University of Cape Town, 2003.
137. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graph.*, 1996, **14**, 33–38.
138. J. E. Stone, *VMD solvate plugin*, http://www.ks.uiuc.edu/Research/vmd/plugins/solvate/.
139. I. Blabin, M. Sotomayor and L. Trabuco, *VMD ionize plugin*, http://www.ks.uiuc.edu/Research/vmd/plugins/autoionize/.
140. M. Kuttel, J. Gain, A. Burger and I. Eborn, *J. Mol. Graph. Model.*, 2006, **25**, 380–388.
141. S. Nosé and M. L. Klein, *Mol. Phys.*, 1983, **50**, 1055–1076.
142. W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
143. S. E. Feller, Y. Zhang, R. W. Pastor and B. R. Brooks, *J. Chem. Phys.*, 1995, **103**, 4613–4621.
144. *NAMD Configuration Files*, 2015, http://www.ks.uiuc.edu/Training/Tutorials/namd/namd-tutorial-win-html/node27.html.
145. A. Shrake and J. Rupley, *J. Mol. Biol.*, 1973, **79**, 351–371.
146. J. E. Stone, *VMD SASA*, 2015, http://www.ks.uiuc.edu/Research/vmd/vmd-1.8.3/ug/node121.html.
147. T. Anderson, *PhD thesis*, University of Cape Town, 2015.
148. A. Tóth, A. Medgyes, I. Bajza, A. Lipták, G. Batta, T. Kontrohr, K. Péterffy and V. Pozsgay, *Bioorg. Med. Chem. Lett.*, 1999, **10**, 19–21.
149. A. Medgyes, E. Farkas, A. Lipták and V. Pozsgay, *Tetrahedron*, 1997, **53**, 4159–4178.
150. H. B. Pfister and L. A. Mulard, *Org. Lett.*, 2014, **16**, 4892–4895.
151. K. Benakli, C. Zha and R. J. Kerns, *J. Am. Chem. Soc.*, 2001, **123**, 9461–9462.
152. B. Guilbert, N. J. Davis, M. Pearce, R. T. Aplin and S. L. Flitsch, *Tetrahedron: Asymmetry*, 1994, **5**, 2163–2178.
153. C. Jones and X. Lemercinier, *J. Pharm. Biomed. Anal.*, 2002, **30**, 1233–1247.
154. P. C. Fusco, E. K. Farley, C. H. Huang, S. Moore and F. Michon, *Clin. Vaccine Immunol.*, 2007, **14**, 577–584.
155. X. Luo, T. Sugiura, R. Nakashima, Y. Kitamura and Y. Kitade, *Bioorg. Med. Chem. Lett.*, 2013, **23**, 4157–4161.